AD-A062 371 WEAPONS RESEARCH ESTABLISHMENT SALISBURY (AUSTRALIA) F/G 14/2
MAXIMUM ENTROPY SPECTRUM ANALYSIS TECHNIQUE - A REVIEW OF ITS T--ETC(U)
APR 77 D A GRAY

UNCLASSIFIED WRE-TR-1812(W) NL

OF 1
AD A062371

END
DATE FILMED
3 -79
DDC


1.0
1.1
4.5
5.0
5.6
6.3
7.1
8.0
9.0
2.8
3.2
3.6
4.0
2.5
2.2
2.0
1.8
1.4
1.6

LEVEL II

WRE-TR-1812 (W)

AR-000-559

AUSTRALIA

DEPARTMENT OF DEFENCE

DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION

WEAPONS RESEARCH ESTABLISHMENT

SALISBURY, SOUTH AUSTRALIA

TECHNICAL REPORT 1812 (W)

MAXIMUM ENTROPY SPECTRUM ANALYSIS TECHNIQUE - A REVIEW OF ITS THEORETICAL PROPERTIES

D.A. GRAY

DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION

Approved for Public Release

COPY No. 25

APRIL 1977

AR-000-559

DEPARTMENT OF DEFENCE

DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION

WEAPONS RESEARCH ESTABLISHMENT

(9) TECHNICAL REPORT 1812 (W)

(6) MAXIMUM ENTROPY SPECTRUM ANALYSIS TECHNIQUE - A REVIEW OF ITS THEORETICAL PROPERTIES.

(10) D.A./Gray

(12) 45 p.

(11) Apr 77

SUMMARY

The maximum entropy method estimates the power spectrum of a random process by extrapolating the autocorrelation function outside the observation interval. The extrapolation is such that a minimal number of constraints is imposed on it. The equivalence of various derivations of the maximum entropy spectrum is proved in this paper and various examples are given to illustrate the properties of the maximum entropy spectrum. Finally the relation of the maximum entropy method to other techniques for estimating the spectrum is given.

(14) WRE-TR-1812 (W)

Approved for Public Release

POSTAL ADDRESS: The Director, Weapons Research Establishment, Box 2151, G.P.O., Adelaide, South Australia, 5001.

DOCUMENT CONTROL DATA SHEET

Security classification of this page: UNCLASSIFIED

1 DOCUMENT NUMBERS	2 SECURITY CLASSIFICATION
AR Number: AR-000-559	a. Complete Document: Unclassified
Report Number: WRE-TR-1812(W)	b. Title in Isolation: Unclassified
Other Numbers:	c. Summary in Isolation: Unclassified

3 TITLE

MAXIMUM ENTROPY SPECTRUM ANALYSIS TECHNIQUE - A REVIEW OF ITS THEORETICAL PROPERTIES

4 PERSONAL AUTHOR(S):	5 DOCUMENT DATE:
D.A. Gray	April 1977
	6 6.1 TOTAL NUMBER OF PAGES 44
	6.2 NUMBER OF REFERENCES: 16

7 7.1 CORPORATE AUTHOR(S):	8 REFERENCE NUMBERS
Weapons Research Establishment	a. Task: DST 20/5
7.2 DOCUMENT (WING) SERIES AND NUMBER Weapons Research and Development Wing TR-1812	b. Sponsoring Agency:
	9 COST CODE: 343587

10 IMPRINT (Publishing establishment):	11 COMPUTER PROGRAM(S) (Title(s) and language(s))
Weapons Research Establishment	

12 RELEASE LIMITATIONS (of the document):

Approved for Public Release

12.0	OVERSEAS	NO		P.R.	1	A		B		C		D		E	

Security classification of this page: UNCLASSIFIED

Security classification of this page: UNCLASSIFIED

13 ANNOUNCEMENT LIMITATIONS (of the information on these pages):

No limitation

14 DESCRIPTORS:

a. EJC Thesaurus Terms: Entropy; Spectrum analysis

b. Non-Thesaurus Terms:

15 COSATI CODES:

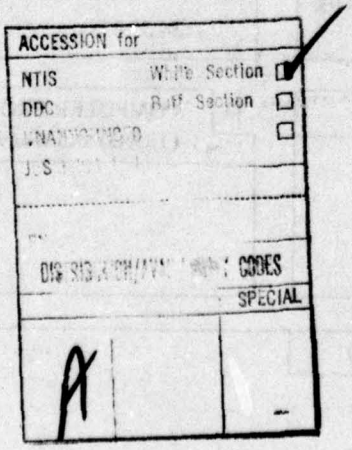
1402

16 LIBRARY LOCATION CODES (for libraries listed in the distribution):

SW SR SD AACA NL

17 SUMMARY OR ABSTRACT:
(if this is security classified, the announcement of this report will be similarly classified)

The maximum entropy method estimates the power spectrum of a random process by extrapolating the autocorrelation function outside the observation interval. The extrapolation is such that a minimal number of constraints is imposed on it. The equivalence of various derivations of the maximum entropy spectrum is proved in this paper and various examples are given to illustrate the properties of the maximum entropy spectrum. Finally the relation of the maximum entropy method to other techniques for estimating the spectrum is given.



Security classification of this page: UNCLASSIFIED

TABLE OF CONTENTS

LIST OF APPENDICES

LIST OF FIGURES

1. INTRODUCTION

In conventional spectrum analysis the resolution of features of interest is often restricted by the finite observation interval. Often, either to enhance particular features, or to improve the statistical reliability of the spectrum, the data are modified by a window function. A proliferation of window functions exist and generally the choice of such a window is a trade-off between resolution and distortion due to side-lobe leakage. A common feature of all windows is the assumption that the autocorrelation function is zero outside the observation interval.

This, in general, is not true and a new approach to the estimation of the power spectrum has been proposed by J.P. Burg(ref.1,2) which extrapolates the autocorrelation function outside the observation interval. Burg's technique, known as the maximum entropy method, extrapolates the autocorrelation function in such a way that a minimal number of constraints is imposed on the extrapolated sequence. This is an application of the general principle of maximum entropy which is discussed in more detail in Section 3. This approach obviously contrasts with the very strict constraint imposed by Fourier analysis that, irrespective of the observed nature of the autocorrelation function within the observation interval, its extension is always zero.

In view of the amount of recent interest in Burg's method this paper has been written as a review of the theory of maximum entropy spectrum analysis. Various derivations of the maximum entropy spectrum are given and a discussion of its properties and its application to some well known random processes are illustrated by numerical examples. No attempt has been made to discuss the application of the method to real data.

2. CONVENTIONAL ESTIMATION OF THE POWER SPECTRUM

Let x_i represent the i-th sample of a discrete random process with an autocorrelation matrix* R defined by $R_{ij} = \langle x_i x_j^* \rangle$. Assuming the process to be ergodic and real the autocorrelation R will be Toeplitz, i.e., $R_{ij} = r_{|i-j|} = \langle x_i x_j \rangle$. The spectrum S(f) of this ergodic process is defined as the Fourier transform of the first row of the matrix R, i.e.:

$$S(f) = \left(\lim_{N\to\infty} \frac{1}{2N+1} \sum_{-N+1}^{N-1} r_\rho \, e^{-2\pi i f \rho \tau_o} \right) \tau_o$$

where τ_o is the time interval between samples.

From a given set of random samples $x_o, x_1, \ldots x_{N-1}$ it is possible to estimate at most N lags of the autocorrelation function. From the above definition of the spectrum of a random process, a natural form for the estimator of the spectrum S(f) would then be

$$\hat{S}(f) = \left(\sum_{-N+1}^{N-1} r_\rho \, e^{-2\pi i f \rho \tau_o} \right) \tau_o$$

where ^ denotes an estimate and so $\hat{r}_\rho$ is some estimate of the ρ-th autocorrelation

*The autocorrelation matrix R defined in this way will in general be of infinite dimension but in this paper the liberty of using either R or R_N to denote the first N rows and N columns will be taken when there is no source of confusion.

lag. As an example, where $\hat{r}_\rho$ is given by

$$\hat{r}_\rho = \frac{1}{N} \sum_{j=0}^{N-|\rho|-1} x_j \, x_{j+|\rho|}$$

the above estimate of the power spectrum can readily be shown to be the well known periodogram. Inherent in the approach however is the assumption that the autocorrelation function r_ρ is zero outside the observation interval. The effect of this truncation is to distort the estimate $\hat{S}(f)$ of the spectrum by convolving it with the transform of a window which in the above case is the triangular one. As a result of this convolution the resulting spectrum is unable to resolve periodic components whose frequencies differ by less than the reciprocal of the observation interval. Also, the estimate of the power at a particular frequency may be distorted due to leakage through sidelobes. For a random process the lack of resolution and sidelobe leakage due to the finite observation period can be interpreted in terms of a variance and bias in the estimate of power spectrum.

In order to reduce either the variance or bias of the estimate of the power spectrum, the autocorrelation function is often multiplied by a window function. Unfortunately, windows which reduce the variance of the power estimate are inferior in resolving closely spaced lines, and, in general, the choice of a window involves a trade-off between resolution and leakage. Typical window functions such as Bartlett, Hanning and Parzen are discussed in reference 3.

An alternative method which both increases the resolution and decreases the leakage is to increase the number of lags of the autocorrelation function by extrapolation. This method does not, as in the case of windowing, distort the observed data. The power spectrum may then be estimated as the transform of the extrapolated sequence. In the remainder of this paper an extrapolation scheme based on the information theoretic concept of entropy will be considered.

3. THE PRINCIPLE OF MAXIMUM ENTROPY APPLIED TO DATA ANALYSIS

3.1 The entropy of a random process

The concept of entropy was originally introduced through the second law of thermodynamics as a measure of the heat lost or adsorbed in the transition of a system from an initial to a final state. The fact that the entropy of an irreversible process always increased, led to the use of entropy as a measure of the disorder of a system, since any irreversible process may be conceptually thought of as increasing the disorder of the system.

A highly disordered or random process is characterized by a lack of constraints, and so, in addition to entropy being a measure of the randomness of a process, it may also be considered as a measure of the constraints (or strictly the lack of constraints) imposed on a system. In particular a highly structured system has had very strong constraints imposed on it. Since a highly structured system contains little information other than that concerning its structure, it follows as a corollary to the above, that entropy may be considered as a measure of the information content of a process. It was in this form that entropy was introduced by Shannon(ref.4) into the theory of random processes, and it forms the basis of modern information theory.

If x_i, the time series samples of a random process, can be considered to form an N-dimensional multivariate random process then Shannon(ref.4) showed that the expectation value of information, i.e., the entropy H is given by

$$H = -\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(\underline{x}) \log p(\underline{x}) \, dx_1 \, dx_2 \cdots dx_N \qquad (1)$$

where $p(\underline{x}) = p(x_1, x_2, \ldots x_N)$ is the joint N-dimensional probability distribution function (p.d.f.). Furthermore, if the x_i's are samples from an N-dimensional Gaussian distribution, the joint p.d.f. is given by

$$p(\underline{x}) = (2\pi)^{-N/2} \{\det R\}^{-\frac{1}{2}} \exp(-\tfrac{1}{2} \underline{x}^t R^{-1} \underline{x})$$

where, as in Section 2, R is defined by $R_{ij} = r_{|i-j|} = \langle x_i x_j \rangle$ for a real process. Substituting the above equation for $p(x)$, equation (1) reduces to

$$H = \log((2\pi e)^{N/2} \{\det R\}^{\frac{1}{2}})$$

where det R represents the determinant of the matrix R.

The fact that R is related to the spectrum S(f) by a Fourier transform relationship suggests that an expression for entropy in terms of the spectrum S(f) of the ergodic process is also possible*. Shannon(ref.4) has shown, however, that only entropy differences are significant for continuous processes. Thus for a Gaussian process the difference in entropy between Gaussian white noise of unit spectral density and an ergodic Gaussian process with spectral power density function S(f) is given by $-\frac{1}{W} |\int_W \log S(f)\, df|$ where W is the bandwidth of the spectrum. It is a well-known fact that Gaussian white noise is the process with maximum entropy and this forms a suitable reference with which to compare entropies. For the remainder of this paper $\frac{1}{W} |\int_W \log S(f)\, df|$ will be termed the entropy associated with random process whose spectrum is S(f). Obviously, maximizing the entropy of a process is equivalent to maximizing $|\int_W \log S(f)\, df|$.

It should be noted that both expressions for the entropy associated with a process can only be defined provided certain conditions (essentially equivalent) are satisfied by either the autocorrelation matrix or the spectrum. In particular $\int_W \log S(f)\, df > -\infty$ or alternatively S(f) should have no zeros. This has a direct physical interpretation in requiring that the process not be deterministic, i.e., entropy in the above form is only applicable to random processes. The effect of this restriction is illustrated in Section 6.4. The equivalent restriction is that the autocorrelation matrix is non-singular and positive definite.

3.2 The principle of maximum entropy

Often the constraints imposed on a random process are insufficient to uniquely determine the process. For example, the mean and variance of a random process do not uniquely determine the probability density function of the process. The principle of maximum entropy is to choose from all the processes which are consistent with the given information the one which maximizes the entropy. Before applying this to spectrum analysis, the reasons for this choice will be elucidated.

In thermodynamics, when a system is in equilibrium its entropy is maximized. Since most physical systems tend towards equilibrium, then the most likely state of the system, given that all other considerations are equal, is the one with the maximum entropy. From this viewpoint, choosing the process which maximizes the entropy would seem to be the most natural choice, since from a

* The fact that det R is just the product of the eigenvalues of R coupled with Szegő's remarkable theorem(ref.5) relating functions of eigenvalues to an integral of the power spectrum, is in fact a stronger inference.

dynamic point of view the system is most likely to be in equilibrium.

An alternative statistical approach when faced with the choice of several competing processes, all of which are consistent with the given information, is to choose the most probable one. An alternative criterion is to choose the process about which the most uncertainty exists. This can be interpreted as choosing the solution with the fewest constraints imposed on it (while still remaining consistent with the available information). Since entropy can be considered as a measure of the lack of constraints imposed on a process the principle of maximum entropy then implies that the process is chosen such that it maximizes the entropy while remaining consistent with the given constraints.

Both approaches give some sort of rationalization for the use of entropy. The maximum entropy principle may be summarized as follows:

(a) It is uniquely determined and is maximally non-committal with respect to missing information; and

(b) It assigns positive weights to every situation not absolutely excluded by the given information.

The principle of maximum entropy is not new and has been applied to fields such as statistical mechanics*(ref.6) and the estimation of probability distribution functions(ref.7,8). It was first applied to power spectrum estimation by Burg(ref.1) who utilized the expressions derived by Shannon for entropy in terms of the autocorrelation function or spectrum. Burg's technique is essentially an extrapolation of the autocorrelation function according to the principle of maximum entropy: that is, the autocorrelation function is estimated outside the observation interval in such a way that the minimal number of assumption is made about this extension. This is in contrast to the conventional approach whereby the very strong and often erroneous assumption is made that the autocorrelation function is zero outside the observation interval.

The concept of a window as discussed in Section 2 does not apply, since the form of the extrapolation depends on the autocorrelation function itself. An alternative approach is to claim that the window is adaptive in that it adjusts itself in accordance with the process.

4. DERIVATIONS OF THE MAXIMUM ENTROPY SPECTRUM

Two alternative derivations of the maximum entropy power spectrum based on the formulae for entropy in Section 3.1 will be given in this Section. The first is based on the extrapolation of the autocorrelation function, while the second directly estimates the power spectrum. Finally the equivalence of the two methods is proved.

4.1 Maximum entropy extrapolation of the autocorrelation function

As discussed in Section 3.1, given the N lags r_0, r_1, ..., r_{N-1} of the autocorrelation function of an ergodic Gaussian process, the entropy is given by

$$H_N = \log((2\pi e)^{N/2} \{\det R_N\}^{1/2})$$

where

*However, in most applications in statistical mechanics the Stirling approximation reduces the maximum entropy and maximum probability formulations to the same thing.

$$R_N = \begin{pmatrix} r_0 & r_1 & \cdot & \cdot & \cdot & r_{N-2} & r_{N-1} \\ \cdot & r_0 & & & & & r_{N-2} \\ \cdot & & \cdot & & & & \cdot \\ \cdot & & & \cdot & & & \cdot \\ \cdot & & & & \cdot & r_1 & \cdot \\ \cdot & & & & & r_0 & r_1 \\ r_{N-1} & & & & \cdot & \cdot & r_0 \end{pmatrix}$$

Let $\tilde{r}_N$ be the extrapolation of this autocorrelation function by one lag. The entropy is now given by

$$H_{N+1} = \log((2\pi e)^{N/2} \{\det \tilde{R}_{N+1}\}^{\frac{1}{2}})$$

where

$$\tilde{R}_{N+1} = \begin{pmatrix} r_0 & r_1 & \cdot & \cdot & \cdot & r_{N-1} & \tilde{r}_N \\ \cdot & r_0 & r_1 & \cdot & \cdot & \cdot & r_{N-1} \\ \cdot & & \cdot & & & & \\ \cdot & & & \cdot & \cdot & \cdot & \\ \tilde{r}_N & \cdot & \cdot & \cdot & \cdot & \cdot & r_0 \end{pmatrix}$$

Van den Bos(ref.9), applying the principle of maximum entropy, has shown that if r_N is chosen, such that H_{N+1} is maximized (i.e., $\partial H_{N+1}/\partial \tilde{r}_N = 0$) then $\tilde{r}_N$ is uniquely obtained as the solution of the determinantal equation

$$\det \begin{bmatrix} r_1 & \cdot & \cdot & \cdot & r_{N-1} & \tilde{r}_N \\ r_0 & r_1 & & & & r_{N-1} \\ \cdot & & \cdot & & & \cdot \\ \cdot & & & \cdot & & \cdot \\ \cdot & & & & \cdot & \cdot \\ r_{N-2} & & & & & r_1 \end{bmatrix} = 0 \qquad (2)$$

If $r_0, r_1, \ldots r_{N-1}, \tilde{r}_N$ is taken as the new correlation sequence, then this sequence may be extended to $\tilde{r}_{N+1}$ by maximizing the entropy associated with the extended sequence. In a similar way $\tilde{r}_{N+2}, \tilde{r}_{N+3}, \ldots$ may be found and the autocorrelation function may be extended indefinitely. Once the autocorrelation function has been extrapolated a sufficient number of lags, the power spectrum may be estimated by taking the Fourier transform of this extended sequence. The power spectrum that results from an infinite extension of the autocorrelation sequence is of particular interest, and a method for obtaining an analytic expression for this will be given in the next section.

As an example of this principle, consider the random process of a sine wave of frequency f in white noise. The autocorrelation lags are given by

$$r_\rho = \delta_{\rho o} + a \cos 2\pi f \rho \tau_o$$

where a is the signal-to-noise ratio and τ_o is the sampling interval. The results of extending by means of the M.E. method 16 lags of this autocorrelation function to 128 lags for a sine wave of frequency 1 Hz is shown in figures 1(b), 1(c) and 1(d) for a = 10, 1 and 0.5 respectively. In this and all following plots of the autocorrelation functions, the white noise component is not plotted and all graphs are normalized to one. The corresponding Fourier extension 1(a) assumes $r_\rho = 0$ for $\rho \geq 16$. All examples used in this paper take $\tau_o = 1/16$ so that the corresponding frequency range is 0 to 8 Hz.

4.2 The maximum entropy power spectrum

Burg has shown how the maximum entropy power spectrum can be obtained directly from the autocorrelation matrix R_N by the evaluation of the associated prediction error filter coefficients (p.e.f.c.'s). His intuitive approach has been rigorously formalized by Edwards and Fitelson(ref.10) and their derivation is outlined here.

As discussed in Section 3 the entropy* associated with a random process can be expressed in terms of the power spectral density by

$$H_\infty = \frac{1}{W} \left| \int_W \log S(f)\, df \right| .$$

If the autocorrelation function is known for lags $\rho = 0,1, \ldots, N-1$ the principle of maximum entropy implies that H_∞ should be maximized subject to the constraint that

$$\int_W S(f)\, e^{2\pi i f \rho \tau_o}\, df = r_\rho$$

By introducing Lagrangian multipliers it can be shown that the S(f) which maximizes H_∞ subject to the above constraints is given by

$$S(f) = \left(\frac{P_N}{A(f)\, A(f)^*} \right) \tau_o \qquad (3(a))$$

where

$$A(f) = \sum_{j=0}^{N-1} \gamma_j\, e^{-2\pi i f j \tau_o} \qquad (3(b))$$

*The entropy is denoted as H_∞ since it corresponds in the sense of Section 4.1 to an infinite extension of the correlation sequence.

P_N is the output power of a prediction error filter and the γ's are the corresponding p.e.f.c.'s. The γ's are related to the Toeplitz autocorrelation matrix by the following equation

$$\begin{pmatrix} r_0 & r_1 & \cdots & r_{N-1} \\ r_1 & r_0 & \cdots & r_{N-2} \\ \vdots & \vdots & \ddots & \vdots \\ r_{N-1} & r_{N-2} & \cdots & r_0 \end{pmatrix} \begin{pmatrix} \gamma_0 \\ \gamma_1 \\ \vdots \\ \gamma_{N-1} \end{pmatrix} = \begin{pmatrix} P_N \\ 0 \\ \vdots \\ 0 \end{pmatrix} \tag{4}$$

where $\gamma_o = 1$.

Figure 2+ shows the M.E. and Fourier spectra corresponding to the autocorrelation lags of figure 1. The M.E. spectra have been calculated according to equations 3 and 4.

4.3 Equivalence of the two approaches

Since the spectrum derived in the previous section is a continuous function of frequency and also periodic it follows that it can be written as the Fourier transform of an infinite correlation sequence, i.e.,

$$S(f) = \left(\sum_{-\infty}^{\infty} r'_\rho \, e^{-2\pi i f \rho \tau_o} \right) \tau_o$$

The r_ρ's are given by the inverse transform, that is

$$r'_\rho = \int_W S(f) \, e^{2\pi i f \rho \tau_o} \, df$$

From the constraint equations 4 it directly follows that $r'_\rho = r_\rho$ for $\rho = 0,1, \ldots, N-1$. A necessary and sufficient condition for the two approaches to be equivalent is that $r'_\rho = \tilde{r}_\rho$ where $\tilde{r}_\rho$'s are the appropriate solutions of the determinantal equation (2).

By an extension of the argument used by Edwards and Fitelson it can easily be shown that equation (4) can be augmented to

+In this and all following graphs the square root of the normalized power spectrum rather than the power spectrum itself is plotted.

$$\begin{pmatrix} r_0 & r_1 & \cdot & \cdot & \cdot & r_{N-1} \\ r_1 & r_0 & r_1 & & & \\ & \cdot & & & & \\ & & \cdot & & & \\ & & & \cdot & & \\ & & & & r_0 & \\ r_{N-1} & & & & r_0 & \\ r'_N & r_{N-1} & & & & \\ \cdot & \cdot & & & \cdot & \\ \cdot & \cdot & & & \cdot & \end{pmatrix} \begin{pmatrix} 1 \\ \gamma_1 \\ \cdot \\ \cdot \\ \cdot \\ \gamma_{N-1} \end{pmatrix} = \begin{pmatrix} P_N \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \\ 0 \\ \cdot \\ \cdot \end{pmatrix}$$

Taking rows 2 to N + 1 this equation becomes

$$\begin{pmatrix} r_1 & r_0 & \cdot & \cdot & \cdot & r_{N-2} \\ r_2 & r_1 & & & & \\ \cdot & & \cdot & & & \\ \cdot & & & \cdot & & \\ r_{N-1} & & & & \cdot & \\ r'_N & r_{N-1} & & & r_1 & \end{pmatrix} \begin{pmatrix} 1 \\ \gamma_1 \\ \cdot \\ \cdot \\ \cdot \\ \gamma_{N-1} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{pmatrix}$$

Since the γ's are unique and non-zero the above equations are only consistent if

$$\det \begin{bmatrix} r_1 & r_0 & \cdot & \cdot & \cdot & r_{N-2} \\ & r_1 & & & & \\ & & \cdot & & & \\ & & & \cdot & & \\ r_{N-1} & & & & \cdot & \\ r'_N & r_{N-1} & & & r_1 & \end{bmatrix} = 0$$

This, however, is exactly the equation derived in Section 4.1 for the M.E. extension $\tilde{r}_N$ of the autocorrelation series. As a consequence of the uniqueness of this solution it follows that $\tilde{r}_N = r'_N$. By an extension of this argument it follows that $\tilde{r}_{N+\rho} = r'_{N+\rho}$ for all ρ.

The significance of the p.e.f.c.'s will be more apparent in later sections, but their use in determining the M.E. extension of the autocorrelation sequence should be pointed out here. Rather than laborously solving equation (2) for each extension of the autocorrelation sequence, it is a much easier task to obtain the p.e.f.c.'s by solving equation (4) and then estimating the ρ-th extension of the autocorrelation sequence by the equation

$$r_{N+\rho} = -\gamma_1 r_{N+\rho-1} - \gamma_2 r_{N+\rho-2} - \cdots - \gamma_{N-1} r_{\rho-1}.$$

This equation is both useful for numerical implementation and for theoretical considerations.

5. PROPERTIES OF THE MAXIMUM ENTROPY SPECTRUM

Many of the important properties of the conventional estimation of the power spectrum can be described either in terms of the linearity of the Fourier transform or related to the window applied, and hence to number of lags of the correlation sequence used. As evidenced by equations 3(a), 3(b) and 4 the M.E. spectrum is not linear. Moreover, as discussed previously, neither the concept of a data window or fixed observation period is applicable to this method. In general in the M.E. method, many of the usual properties of the power spectrum such as positivity, linearity, etc., will depend on the nature of the autocorrelation function. A number of the following properties have not been rigorously proved but have been inferred from the results of computer simulation.

5.1 Existence and positivity

In conventional analysis, provided the autocorrelation function is non-negative definite, a positive spectrum will always exist. In order to form the M.E. spectrum, and, in particular, to solve equation (4), the autocorrelation matrix must also be non-singular. This then implies that the roots of $A(z)$, where $z = e^{2\pi i \tau_o}$, lie within the unit circle and so the resulting maximum entropy spectrum, $S(z) = {}^{1}/A(z)\, A^*(z)$ will be stable. It also follows from equation (3) that $S(z)$ is an all-pole model. This has an important and interesting consequence. If the true spectrum of a process has a power spectrum that is zero over any finite frequency interval then the M.E. technique, since it corresponds to an all-pole model, will not be able to approximate these zeros unless a large number of poles are present. An alternative method would be to introduce a small amount of white noise into the system and so make the spectrum positive. This is equivalent to making the autocorrelation matrix non-singular. Why the method breaks down can easily be seen from the expression for entropy $\frac{1}{W} \left| \int_W \log S(f)\, df \right|$. If $S(f)$ is zero over any finite region then this expression is unbounded and the formulation breaks down. The restriction that $\int_W \log S(f)\, df > -\infty$ is known as the Paley-Wiener criterion and is usually interpreted as implying that the physical process is non-deterministic. This immediately leads to the surprising conclusion that the M.E. technique as formulated in terms of an all-poles model is only applicable to random (and not deterministic) processes. This again is in contrast to conventional analysis where the autocorrelation approach to power spectrum estimation applies to both random and non-random processes.

The positivity of the M.E. spectrum is guaranteed by equation (3) where, provided R is positive definite, then P_N is greater than zero.

5.2 Power and frequency estimation

In conventional analysis, it is often asserted that the power of a particular frequency component is proportional to the height of that spectral line if the contributions due to other signals and noise are ignored. It is also true under the same conditions that if the area within the main peak and all its sidelobes is integrated, then the power is proportional to this area. If the contribution due to sidelobes is neglected then the power is proportional to the area under the main lobe. In general, the height of a spectral line in the M.E. method is not proportional to the power of that spectral component, and, indeed, Lacoss has shown that this height can have a large variance associated with it. He has however shown that the area under the peaks is approximately proportional to the power of that spectral component.

The frequency of a spectral component is, in both the conventional and M.E. method, determined by the position of peaks in the power spectrum. At

low frequencies these peaks are often "shifted" due to the interference of positive and negative sidelobes in conventional analysis. This is particularly illustrated in figure 6(a) where for N = 8 the peak of the spectrum is somewhat greater than 2 Hz. Due to the lower sidelobe level the corresponding M.E. spectrum (figure 6(b)) does not exhibit the effect.

5.3 Resolution

In conventional analysis, signals differing in frequency by $\Delta f < \frac{1}{T}$ are not considered resolvable (the Rayleigh criterion or uncertainty principle). The effect of a window will in general only reduce the sidelobes at the expense of decreasing the resolution. Since the M.E. technique can effectively extend the data window indefinitely, the increase in resolution of the M.E. method will depend on how faithfully the autocorrelation function is extrapolated. The M.E. extrapolations of 16 lags of the autocorrelation function of a 1 Hz sine wave in white noise for a = 10, 1 and 0.5 are shown in figures 1(b), 1(c) and 1(d) respectively. Taking 32, 16 and 8 lags of the same autocorrelation function with a = 1, the M.E. extensions of the autocorrelation functions are shown in figures 3(b), 3(c) and 3(d) respectively. These figures indicate that the M.E. extrapolation is modulated by an exponential decay with a decay rate inversely proportional to the number of lags and to the signal-to-noise ratio. That this modulation is independent of frequency is shown by figures 4 and 5 which differ only from figure 1 in that the frequency of the sine wave is 2 Hz and 4 Hz respectively.

In figure 2 the spectra corresponding to the autocorrelation sequences in figure 1 are plotted. The improvement of the spectra in the form of a decreased sidelobe level and line width is a consequence of the improvement in the M.E. extrapolation as the S.N.R. is increased. As the number of lags is increased the M.E. spectrum will, in a similar manner to the Fourier case, be characterized by a lower average sidelobe level and narrower line width. To illustrate this the spectra corresponding to the autocorrelation sequences in figure 3 are plotted in figure 6.

In order to demonstrate the superior ability of the M.E. technique to resolve closely spaced lines, the M.E. spectra for two sine waves in white noise were calculated. For such a process the autocorrelation function is given by

$$r_\rho = \delta_{\rho o} + a_1 \cos 2\pi f_1 \rho\tau_o + a_2 \cos 2\pi f_2 \rho\tau_o \qquad (5)$$

where a_1, a_2, f_1 and f_2 are the S.N.R.'s and frequencies of the respective sine waves.

In figure 7, 16 lags of the autocorrelation function are used to estimate the Fourier and M.E. spectra of two sine waves of frequencies 3 and 3.5 Hz in white noise. The S.N.R.'s, a_1 and a_2 are taken to be equal in figure 7. The superiority of the M.E. technique is clearly demonstrated for high S.N.R.'s but as the S.N.R.'s are decreased the M.E. estimate of the power spectrum approaches the Fourier one.

The dependence of the resolution on the number of lags is shown in figure 8 where the spectra resulting from using 4, 8 and 16 lags of the above autocorrelation function are shown with $a_1 = a_2 = 20$ and $f_1 = 3$ and $f_2 = 3.5$ Hz. Once again the resolving power of the M.E. method is superior to the Fourier case with this superiority increasing as the number of lags is increased. It should be noted that although the M.E. technique is able to resolve the two signals its estimate of the relative amplitudes (see figure 8(b)) may be incorrect.

In some applications it may be important to determine, given a certain number of lags of the autocorrelation function, how many signals are resolvable. In the case of conventional Fourier analysis it is well known

that for N lags of the autocorrelation function then the limit of resolution is N/2 independent spectral lines each separated by $\frac{1}{T}$ Hz where $T = N\tau_o$. A modified version of the theorem holds for maximum entropy spectrum analysis. As discussed by van den Bos, the M.E. spectrum is equivalent to a least-squares all-pole fit to the data. The number of poles, or alternatively the number of frequency components, is thus equal to the number of independent zeros of the polynomial $\gamma_o + \gamma_1 Z + \ldots + \gamma_{N-1} z^{N-1}$. As in the case of conventional analysis due to the reality of the γ's this number is N/2 and so only N/2 independent frequency components are resolvable. This is illustrated in figure 9 where both the Fourier and M.E. techniques are unable to resolve the 10 different frequency components using only 8 lags of the autocorrelation function

$$r_\rho = \delta_{\rho o} + a \sum_{j=1}^{10} \cos 2\pi f_j \rho \tau_o$$

and where the amplitudes of the 10 sine waves are taken to be equal and f_j, the corresponding frequencies, are 0, 0.5 Hz, ..., 4.5 Hz. In this figure, in contrast to previous figures, an increase in a (an effective S.N.R.) does not enable the 10 sine waves to be resolved.

However the advantage of the M.E. technique is that provided the number of lines is less than or equal to N/2 then their resolvable frequency separation is not $\frac{1}{T}$ as in conventional analysis but is determined, as would be expected, by the S.N.R. This is illustrated in figure 10 where 16 lags of the autocorrelation function, r_ρ, are used to estimate the spectrum of 5 equal amplitude sine waves in white noise of frequencies 0.75 Hz, 1.5 Hz, 2.25 Hz, 3 Hz and 3.75 Hz. As illustrated the Fourier technique is unable to resolve the sine waves for all values of a whereas the M.E. technique, although once again giving poor estimates of the relative amplitudes, is able to resolve the components as a is increased. The autocorrelations corresponding to figure 10 are shown in figure 11 where, as in previous figures, the white noise component of the autocorrelation has been suppressed.

5.4 Power leakage and sidelobes

The distortion of power spectrum estimates by the leakage of the power of a strong signal through sidelobes (or ripple) is well known. Sidelobes, although of different shape, are also present in the M.E. spectrum. In contrast to conventional analysis the height of these sidelobes depends, as does the resolution, on how effectively the maximum entropy method extends the autocorrelation function. This in turn depends on the S.N.R. and in figures 2(b) to 2(d) it can be seen that the sidelobe level is reduced as the S.N.R. is increased. It follows that for the M.E. method, stronger signals are better able to be resolved and are less likely to distort weaker signals due to the leakage of its power through sidelobes. This effect is well illustrated in figures 12 and 13 where the spectra correspond to the autocorrelation function of equation (5) with the 3 Hz frequency component 10 dB lower in power than the 3.5 Hz component. In conventional analysis, for all values of a_1, the sidelobes of the stronger signal "swamp" the peak of the weaker signal so as to make it difficult to detect. As shown by figures 12(b) to 12(d), as a_1 is increased the M.E. technique is able to resolve the two signals, but for small values of a_1 the M.E. spectrum approaches the Fourier one. As would be expected, increasing the number of autocorrelation

lags also increases the ability of the M.E. method to resolve the signals, and this is illustrated in figure 13 for 4, 8 and 16 lags of the autocorrelation function with $a_1 = 20$.

5.5 Linearity

If the autocorrelation matrix R equals $R_a + R_b$ then by the linearity of the Fourier transform it follows that the resulting Fourier spectrum, S(f), equals $S_a(f) + S_b(f)$ where $S_a(f)$ and $S_b(f)$ are the spectra corresponding to R_a and R_b respectively. Due to the highly non-linear relationship of the M.E. spectrum to the autocorrelation function, such a property will not in general hold true. Lacoss(ref.11) has made the general statement that this linearity does hold near peaks in the power spectrum. His observation is based on the study of two sine waves in white noise. To illustrate this, the M.E. spectrum of two sine waves in white noise have been computed and are compared with the M.E. spectra of the two separate processes in figures 14 to 16. The upper graphs are the linear combination of the spectra corresponding to the autocorrelation functions

$$r_\rho^{(a)} = \tfrac{1}{2}\,\delta_{\rho o} + a_1 \cos 2\pi f_a \rho\tau_o$$

and

$$r_\rho^{(b)} = \tfrac{1}{2}\,\delta_{\rho o} + a_2 \cos 2\pi f_b \rho\tau_o$$

The lower graphs are derived from the autocorrelation function

$$r_\rho = \delta_{\rho o} + a_1 \cos 2\pi\, f_a \rho\tau_o + a_2 \cos 2\pi\, f_b\, \rho\tau_o$$

which is the linear combination of $r_\rho^{(a)}$ and $r_\rho^{(b)}$. Figures 14 and 15, corresponding to frequencies 2 and 2.5 Hz and 2 and 2.25 Hz respectively show that, for two equal amplitude sine waves, the approximation of linearity holds best for large values of the S.N.R.'s. This is also illustrated in figure 16 where the amplitude of the 2.5 Hz line is one-tenth that of the 2 Hz sine wave. In general it appears that the presence of a second signal will increase both the line width and sidelobe level of the "spectrum" associated with the first and vice versa. In particular in figure 15(c) this broadening is so great that the two spectral lines are no longer resolvable.

5.6 Bias and variance

In estimating any ensemble property of a random process, two important quantities to be minimized are the bias and the variance of that estimate. For conventional estimation of the spectrum these tend to be competing processes, that is, techniques such as windowing will reduce the variance of the estimate at the expense of increasing the bias. A general figure of merit which is accepted as measuring the "goodness" of a spectrum estimator is the equivalent number of degrees of freedom. From this, confidence intervals may be estimated. Such expressions do not appear to have been calculated for the M.E. method. It has been shown(ref.12) that the M.E. power spectrum is asymptotically unbiased (consistent) - i.e., as the number of lags approaches infinity the bias approaches zero. This, however, is hardly the case of interest since the M.E. method is of greatest use when the number of available lags is small.

Perhaps the most significant work has been effected by Lacoss who experimentally introduced statistical fluctuations into the theoretical autocorrelation function of a sine wave in white noise. His conclusions were that while the height of the spectral peak may fluctuate rapidly, i.e., have a large variance, fluctuations of the area under the peak did not significantly differ from those associated with conventional analysis.

6. NUMERICAL EXAMPLES OF THE MAXIMUM ENTROPY SPECTRUM

As an illustration of the method, the M.E. spectrum have been calculated numerically for a number of random processes. The autocorrelation function for these processes can be derived analytically and subsequently used to calculate the M.E. spectrum. The main steps in such calculations were the solution of the normal equations (equation (4)) and the estimation of the power spectrum of the prediction error filter coefficients. An efficient method for the solution of the normal equations was used and is discussed in Appendix I. The power spectrum of the prediction error filter coefficients was evaluated using the FFT (Fast Fourier Transform) routine where the technique of zero filling was used to obtain an arbitrarily fine resolution. It should be emphasized that the technique of zero filling is an exact evaluation of the M.E. spectrum and not an interpolation as it is in the case of conventional analysis.

The following are the results of some typical processes studied.

6.1 A sine wave in white noise

The autocorrelation function of a sine wave in white noise is given by

$$r_\rho = \delta_{\rho o} + a \cos 2\pi\rho\tau_o$$

where a is the signal-to-noise ratio (SNR)*. Both the M.E. spectrum and the M.E. extension of the autocorrelation sequence of this process have been extensively discussed and compared with the conventional estimates in the preceding sections. The results are displayed in figures 1 to 6. For a high SNR the effective data window can be many times that of the truncated sequence and consequently the M.E. spectrum is characterized by fine resolution and low sidelobes (ripple). As the SNR is decreased, the extrapolated lags of the autocorrelation function approach zero more quickly and the corresponding M.E. spectrum as illustrated by figures 2 and 6 approaches the conventional Fourier one. Increasing the number of lags used for a given SNR also increases the effective range of the extrapolation as shown in figure 3.

6.2 An arbitrary number of sine waves in white noise

The autocorrelation function of two uncorrelated sine waves of frequencies f_1 and f_2 in white noise is given by

$$r_\rho = \delta_{\rho o} + a_1 \cos 2\pi f_1\rho\tau_o + a_2 \cos 2\pi f_2\rho\tau_o$$

where a_1 and a_2 are the respective SNR's.

In figures 7 and 8 the SNR's are taken to be the same and the frequency separation is one-half that determined by the Rayleigh criterion. For all values of the SNR's, conventional analysis fails to resolve the two components. As shown by figures 7(b) to 7(c), as the SNR is increased the two frequencies become resolvable when the M.E. technique is applied.

*As discussed earlier, in all examples $\tau_o = \frac{1}{16}$ s

As a second example, consider the case above with the higher frequency 10 dB in power weaker than the lower frequency component. Once again for all values of the SNR in conventional analysis the higher frequency component is completely swampled by the sidelobes of the lower frequency component. However for the M.E. method as the SNR's are increased (while still keeping the same 10 dB relative difference) the lower power component becomes discernible. This is illustrated in figure 12. The above results lead to the conclusion that the ability of the M.E. method* to resolve two signals is a function of their SNR's.

6.3 Butterworth spectra

For a two-level signal $x(t) = \pm a$ where the duration of each stage is a random variable with an exponential distribution function, the autocorrelation lags r_ρ are given by

$$r_\rho = a^2 e^{-2f_o|\rho|\tau_o}.$$

Figure 17 is a comparison of the Fourier and M.E. spectra for N (the number of lags) equal to 2, 4 and 8. These illustrate a result, proved in Appendix II, that provided more than two lags of the autocorrelation function are used, the M.E. extrapolation in this case becomes the exact one. This is particularly illustrated in figure 17 for $f_o = 1.5$ Hz where apart from an aliasing effect due to the sampling interval the M.E. spectrum is identical to the exact power spectrum. The M.E. extrapolation of the autocorrelation for all cases is compared in figure 18(a) with the corresponding Fourier extrapolation in figures 18(b) to 18(d).

6.4 Band-limited white noise

Pure band-limited white noise has a zero power spectral density over some finite frequency interval and so by the Paley-Wiener criteria is a deterministic process. As discussed in chapter 4, the M.E. method is not applicable to the case, but in a similar fashion to the treatment of a sine wave this can be overcome by adding broadband white noise. The autocorrelation function is then given by

$$r_\rho = \frac{2}{\pi\rho\tau_o} \sin \pi\rho\tau_o(f_H - f_L) \cos \pi\rho\tau_o(f_H + f_L) + N_o \delta_{\rho o}$$

where f_L and f_H are low and high frequency cutoffs respectively of the white noise, τ_o is the sampling interval and N_o is the relative power of the broadband white noise which has been added.

For all values of N_o the roll off of the M.E. spectrum is steeper than the corresponding Fourier one and the ripple outside the bandwidth is lower. However the ripple within the bandwidth is considerably higher and as N_o is decreased this ripple increases considerably until the method eventually breaks down. This is illustrated in figure 19(b) to 19(d) as No is decreased from 10 to 0.1.

*The ability of the M.E. method to resolve a large number of sine waves is shown in figures 9 to 11. The maximum number of sine waves that are resolvable is related to the order of the all poles model that has been used. This has been extensively discussed in Section 5.4.

In general, these examples show that for a stationary random process the M.E. technique of spectrum analysis is superior to the conventional Fourier one. It should be stressed however that these observations are based on the study of simulated ensemble correlation functions. In practice we do not have such a correlation function and must use an estimate of the autocorrelation function which has been obtained from the observed data. The estimation of this autocorrelation function from the data is of critical importance in determining the success of the M.E. method. Techniques and problems associated with this will be addressed in a later paper. It should suffice to mention an interesting technique due to Burg(ref.2) whereby the p.e.f.c.'s can be estimated directly from the data without the need for estimating the autocorrelation function.

7. RELATIONSHIP OF THE MAXIMUM ENTROPY METHOD TO OTHER TECHNIQUES OF SPECTRUM ESTIMATION

The relationship of the M.E. method to the theories of Maximum Likelihood spectrum estimation, prewhitening and linear prediction filtering will be briefly considered. Some results comparing the maximum entropy and the maximum likelihood techniques for the estimation of the power spectra of a sine wave in white noise will be given.

7.1 Autoregressive processes and all-pole models

An autoregressive process of order p-1 is a model where at each instant of time the sampled value of the process x_t is deterministically related to its previous p-1 values plus an additive white noise component. The equation determining this relationship takes the form

$$x_t = a_1 x_{t-1} + a_2 x_{t-2} + \ldots + a_{p-1} x_{t-p+1} + \xi_t \qquad (6)$$

where ξ_t are uncorrelated noise of zero mean satisfying $\langle \xi_t \xi_k \rangle = \sigma \delta_{tk}$ where σ is the noise power. Assuming that equation (6) holds for all t it can be easily shown that

$$\begin{pmatrix} r_0 & r_1 & \cdots & r_{p-1} \\ r_1 & r_0 & & \\ \vdots & & \ddots & \\ r_{p-1} & & \cdots & r_0 \end{pmatrix} \begin{pmatrix} 1 \\ -a_1 \\ \\ -a_{p-1} \end{pmatrix} = \begin{pmatrix} \sigma \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

where $r_\rho = \langle x_t \, x_{t+\rho} \rangle$.

The autoregressive spectrum S(f) is consequently given by

$$S(f) = \frac{\sigma \tau_0}{\left| 1 - \sum_{\rho=1}^{p-1} a_\rho \, e^{-2\pi i f \rho \tau_0} \right|^2}$$

It can easily be seen from a comparison of the above equations with equations (3) and (4) that the M.E. spectrum is just an autoregressive process of order the number of lags used. It follows that any techniques, such as maximum likelihood techniques, for estimating the a_j's directly from the data can be used. An autoregressive process is an all-pole model and extensive use has been made of this in chapter 6 in determining the stability requirements.

7.2 Prewhitening and linear prediction

For a one-step linear prediction filter of order p-1, the j-th value of x is estimated as a linear combination of the previous p-1 values, i.e.,

$$\hat{x}_j = \sum_{k=1}^{p-1} g_k \, x_{j-k}.$$

Provided the g's are chosen to minimize

$$\| x_j - \hat{x}_j \|^2$$

in a least mean square sense then the g's can be shown to be determined by the equation

$$\begin{pmatrix} r_0 & r_1 & \cdots & r_{p-1} \\ r_1 & r_0 & & \\ \vdots & & \ddots & \\ r_{p-1} & & & r_0 \end{pmatrix} \begin{pmatrix} g_1 \\ g_2 \\ \vdots \\ g_{p-1} \end{pmatrix} = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_p \end{pmatrix}$$

Rearranging the above equation the g's can be related to the prediction error filter coefficients in the following way

$$\gamma_i = -g_i \text{ for } i = 1, \ldots, p-1$$

and

$$P_p = r_o - \sum_{k=1}^{p-1} g_k \, r_k.$$

This result has an important consequence. In the previous sections the principle of maximum entropy has been used to extrapolate the autocorrelation function. No reference has been made as to how to apply this technique directly to the data. The relationship between the γ_j's and the g_j's provides such a method for extrapolating the data directly. A greater discussion of this and its application to filtering is given in reference 13.

It can also be shown that the γ_i's are the least mean square solution of the filter equation

$$\underset{\sim}{x} * \underset{\sim}{\gamma} = \underset{\sim}{\delta}$$

where * denotes convolution and the $\underset{\sim}{\delta}$ represents a white noise process. In this sense the M.E. spectrum is related to the filter designed to pre-whiten the x's in an optimal manner.

7.3 Maximum likelihood spectrum analysis

The technique for estimating the maximum likelihood spectrum has a simple physical interpretation. A filter is designed that a sine wave of frequency f is passed undistorted (i.e., no bias) while all other frequencies are rejected in an optimal manner. The optimal manner in this case is least mean square sense (i.e., minimum variance). Applying this to all frequencies the maximum likelihood estimator of the spectrum is given by the power output of this filter. Denoting this spectrum as $S_{ML}^{(N)}(f)$ it can be shown to be given by

$$S_{ML}^{(N)}(f) = \frac{\tau_o}{\sum_{k,l=0}^{N-1} e^{2\pi i f k \tau_o} R_{kl}^{-1} e^{-2\pi i f l \tau_o}}$$

where for stationary processes R_{ij}^{-1} is the inverse of the N x N autocorrelation matrix $R_{ij} = r_{|i-j|}$.

The relationship between the maximum entropy spectrum and the maximum likelihood spectrum is given in the following way. Denote by $S_{ME}^{(p+1)}(f)$ the maximum entropy spectrum obtained from using only p lags of the autocorrelation function and allow p = 0,1, ..., N-1. By exploiting the Toeplitz nature of the autocorrelation matrix Burg(ref.14) has derived the following equation

$$\frac{1}{S_{ML}^{(N)}(f)} = \frac{1}{N} \sum_{p=1}^{N} \frac{1}{S_{ME}^{(p)}(f)} \qquad (7)$$

In an analogy with electrical circuits it would be expected that the $S_{ML}^{N}(f)$ would through the "interference" effect on the R.H.S. of the above equation, have lower ripple in the sidelobes while its main beam would be broader.

Figures 20 and 21 are a comparison of the corresponding maximum entropy and maximum likelihood spectra for a sine wave in white noise. The narrower mainlobe of the maximum entropy and its correspondingly lower sidelobes follow from equation (7). The ripple on the sidelobes is considerably reduced in the maximum likelihood case as would be expected from equation (7).

In terms of resolving power, the maximum entropy technique appears to be superior to the maximum likelihood method. This is illustrated in figure 22

where the former technique will resolve two closely spaced sine waves of equal amplitude and frequencies 3 and 3.5 Hz but the latter will not. In figure 23 the superiority of the Maximum Entropy method in resolving these sine waves when their relative amplitudes differ by 10 dB is shown. Both techniques show an increase in resolving power as the S.N.R. is increased.

8. CONCLUSIONS

Whilst there are a number of theoretical problems still to be solved, the maximum entropy technique can offer considerable advantages over the conventional Fourier techniques. From the studies presented of known random processes, this is particularly true when the signal-to-noise ratio is high. In this sense the maximum entropy technique is perhaps better viewed as an estimator rather than a detector.

Applying the techniques presented here to real data would first involve the estimation of the autocorrelation function from the data. For the conclusions of the studies of simulated correlation sequences to be valid, it is necessary that a "good" statistical estimate of the autocorrelation function be obtained. Provided such care is taken then the application of the technique to real data can be expected to show considerable benefits.

9. ACKNOWLEDGEMENTS

The author is indebted to Dr R.K. Jarrott for useful discussions on the topics covered in this paper. Messrs G.C. Mountford and H.A. d'Assumpcao have contributed useful suggestions during this investigation.

REFERENCES

No.	Author	Title
1	Burg, J.P.	"Maximum Entropy Spectrum Analysis". Paper presented at 37th meeting of the Society of Exploration Geophysicists, Oklahoma City, Oklahoma
2	Burg, J.P.	"A New Analysis Technique for Time Series Data". Paper presented at NATO Advanced Study Institute on Signal Processing, August 1968, Enschede, Netherlands
3	Parzen, E.M.	"Mathematical Considerations in the Estimation of Spectra". Technometrics 3, 167 (1961)
4	Shannon, C.E. and Weaver, W.	"The Mathematical Theory of Communication". Urbana, I11: Univ. of Illinois Press
5	Grenander, V. and Szegō, G.	"Toeplitz Forms and their Applications". University of California Press (1958)
6	Jaynes, E.T.	"Information Theory and Statistical Mechanics". Physical Review Vol. 106, 620 (1957)
7	Good, I.J.	"Maximum Entropy Hypothesis Formulation Especially for Multidimensional Contingency Tables". Annals of Mathematical Statistics Volume 34, 911 (1963)
8	Dawson, D.C. and Wragg, A.	"Maximum-Entropy Distributions Having Prescribed First and Second Moments". IEEE Trans. Info. Theory, 689 (1973)
9	van den Bos, A.	"Alternative Interpretation of Maximum Entropy Spectral Analysis". IEEE Trans. Info. Theory, 493 (1971)
10	Edwards, J.A. and Fitelson, M.M.	"Notes on Maximum-Entropy Processing". IEEE Trans. Info. Theory, 232 (1973)
11	Lacoss, R.T.	"Data Adaptive Spectral Analysis Methods". Geophysics, 36, 661 (1971)
12	Kromer, R.E.	"Asymptotic properties of the Autoregressive Spectral Estimator". (Dissertation) Tech. Report 13 on Contra Nonr-225-(80), Stanford University
13	Ulrych, T.J., Smylie, D.E., Jensen, O.G. and Clarke, G.K.C.	"Predictive Filtering and Smoothing of Short Records by using Maximum Entropy". Jnl. Geophys. Res., 78, 4959 (1973)
14	Burg, J.P.	"The Relationship between Maximum Entropy Spectra and Maximum Likelihood Spectra". Geophysics, 37, 375 (1972)

No.	Author	Title
15	Markel, J.D. and Gray, A.H. Jr.	"On Autocorrelation Equations as Applied to Speech Analysis". IEEE Trans. Audio and Electroacoustics, AO-21, 69 (1973)
16	Bateman, A.	"Higher Transcendental Functions". Vol.2, Ed. Erdelyi, Magnus, Oberhettinger, Tricomi (McGraw-Hill, N.Y. 1953) p. 186

APPENDIX I

SOLUTION OF THE NORMAL EQUATIONS

The Robinson-Levinson algorithm(ref.15) is a computationally efficient method for solving the normal equation

$$\begin{pmatrix} r_0 & r_1 & \cdots & r_{N-1} \\ r_1 & r_0 & & \\ \vdots & & \ddots & \\ & & r_1 & r_0 \end{pmatrix} \begin{pmatrix} 1 \\ \gamma_1 \\ \vdots \\ \gamma_{N-1} \end{pmatrix} = \begin{pmatrix} P_N \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

The method is based on a recursive technique which works directly with the γ's and the P_N and obviates the necessity of inverting the matrix.

Denote by R_p the Toeplitz matrix derived from R by taking the first p rows and columns, and let $\underset{\sim}{\gamma}^{(p)}$ and P_p be the solution of the corresponding normal equation, i.e.

$$R_p \, \underset{\sim}{\gamma}^{(p)} = P_p \, \underset{\sim}{\delta}^{(p)}$$

where

$$\underset{\sim}{\delta}^{(p)t} = (1,0,0,\ldots 0).$$

An auxiliary (p x p) matrix E_p is defined as

$$E_p = \begin{pmatrix} 0 & 0 & \cdots & 0 & 1 \\ 0 & & & 1 & 0 \\ & & \iddots & & \\ 0 & 1 & \iddots & & \\ 1 & 0 & \cdots & \cdots & 0 \end{pmatrix}$$

and can easily be shown to satisfy

$$E_p^2 = I$$

and (I.1)

$$E_p \, R_p \, E_p = R_p.$$

Augmenting the vector $\underset{\sim}{\gamma}^{p}$ by zero in the $(p+1)^{th}$ row and denoting this vector by $\underset{\sim}{a}^{(p+1)}$ it can easily be shown that

$$R_{p+1}\,\underset{\sim}{a}^{(p+1)} = P_p\,\underset{\sim}{\delta}^{(p+1)} + C_p\,E_{p+1}\,\underset{\sim}{\delta}^{(p+1)} \tag{I.2}$$

where

$$C_p = r_p + r_{p-1}\,\gamma_1 + \ldots + r_1\,\gamma_{p-1}. \tag{I.3}$$

Using equations (I.1) it also follows that

$$R_{p+1}\,E_{p+1}\,\underset{\sim}{a}^{(p+1)} = P_p\,E_{p+1}\,\underset{\sim}{\delta}^{(p+1)} + C_p\,\underset{\sim}{\delta}^{(p+1)} \tag{I.4}$$

By virtue of equations (I.2) and (I.3), it follows that the solutions for the $(p+1)^{th}$ order normal equations $\underset{\sim}{\gamma}^{(p+1)}$ and P_{p+1} can be obtained from following recursive relationships

$$\underset{\sim}{\gamma}^{(p+1)} = \underset{\sim}{a}^{(p+1)} - C_{p/P_p}\,E_{p+1}\,\underset{\sim}{a}^{(p+1)} \tag{I.5}$$

and

$$P_{p+1} = P_p - C_p^2/P_p. \tag{I.6}$$

Evaluating equations (I.4), (I.5) and (I.6) for p = 0,1,...,N-1, the solution of the normal equation is obtained.

Comments

(1) A computational advantage of this method over the normal method of inverting the matrix is that the storage is reduced from $\sim N^2$ to $\sim N$ and time is further reduced from $\sim N^3$ to $\sim N^2$.

(2) Unfortunately the recursive technique is susceptible to round-off errors particularly for large N. For a more detailed discussion of this, the reader is referred to reference 15.

(3) The recursive technique may be extended to the solution of an arbitrary equation and hence to the inversion of the matrix. This approach was used in the evaluation of the Maximum Likelihood Spectrum in Section 7.3.

APPENDIX II

THE MAXIMUM ENTROPY ESTIMATE OF THE BUTTERWORTH POWER SPECTRUM

For the random process discussed in 6.3, the autocorrelation function is given by

$$r_\rho = a^2 \; e^{-2\pi f_o |\rho| \tau_o}.$$

Taking, without any loss of generality, $a = \pm 1$ and defining $z = e^{-2\pi f_o \tau_o}$ the autocorrelation matrix can be written

$$R = \begin{pmatrix} 1 & z & z^2 & \cdots & z^{N-1} \\ z & 1 & \cdot & & \\ & \cdot & \cdot & & \\ & & & \cdot & \\ & & & z & 1 \end{pmatrix}.$$

It can readily be verified that the inverse of this is given by

$$R^{-1} = \frac{1}{1 - z^2} \begin{bmatrix} 1 & -z & 0 & \cdots & 0 \\ -z & 1+z^2 & -z & & \cdot \\ 0 & -z & 1+z^2 & & \cdot \\ \cdot & & & \cdot & \\ \cdot & & & 1+z^2 & -z \\ \cdot & & & -z & 1 \end{bmatrix}$$

and hence the p.e.f.c.'s are given by

$$\underset{\sim}{\gamma} = \begin{pmatrix} 1 \\ -z \\ 0 \\ \cdot \\ \cdot \\ \cdot \\ 0 \end{pmatrix} \text{ and } P_N = 1 - z^2.$$

It then follows that the Maximum Entropy spectrum $S_{M.E.}(f)$ is given by

$$S_{M.E.}(f) = \frac{(1 - z^2)\,\tau_o}{1 + z^2 - 2z \cos 2\pi f \tau_o}$$

Since the γ_i's are zero for $i \geq 2$ it follows that only two lags of the autocorrelation function are needed to estimate the M.E. spectrum. Provided

$|z| < 1$ use can be made of the generating function for Tchebyshev polynomials (ref.16) to show that

$$S_{M.E.}(f) = \left(1 + 2 \sum_{j=1}^{\infty} z^j \cos 2\pi f_j \tau_o\right) \tau_o$$

From this it follows that for $N \geq 2$ the Maximum Entropy extrapolation is the exact extrapolation.

The above example can readily be seen to be a first-order autoregressive process and is an example of the general result that if the process is an autoregressive process of order N then the corresponding Maximum Entropy Spectrum is the exact spectrum.

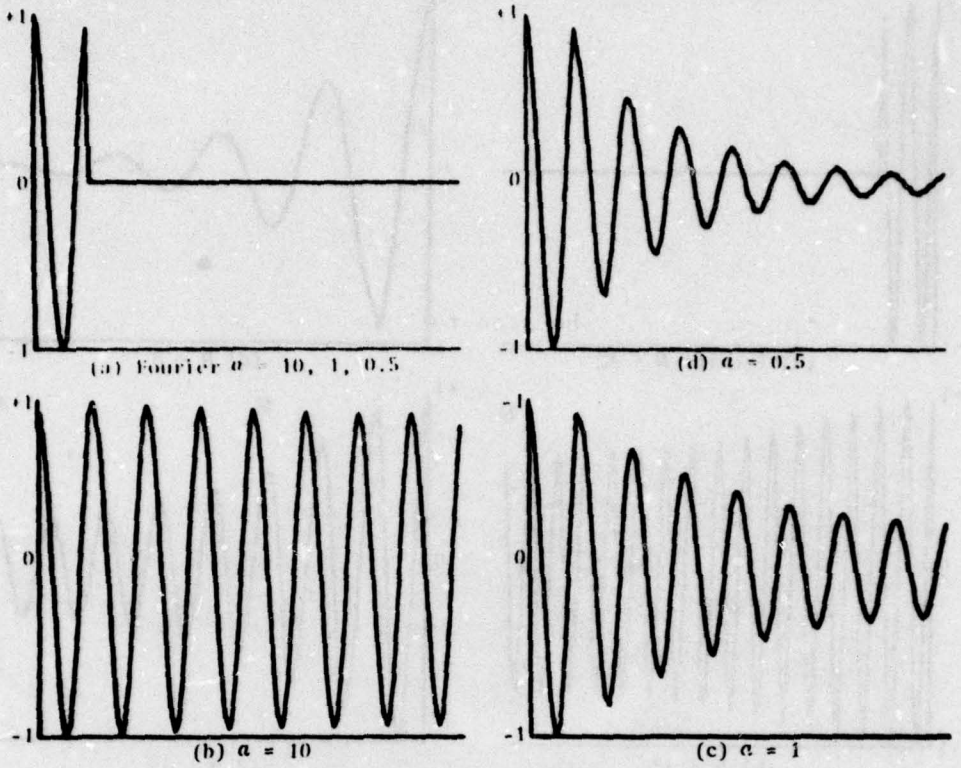


Figure 1. Extrapolation of sine wave in white noise for differing S.N.R.'s
(A: Fourier, B,C,D: Maximum Entropy)

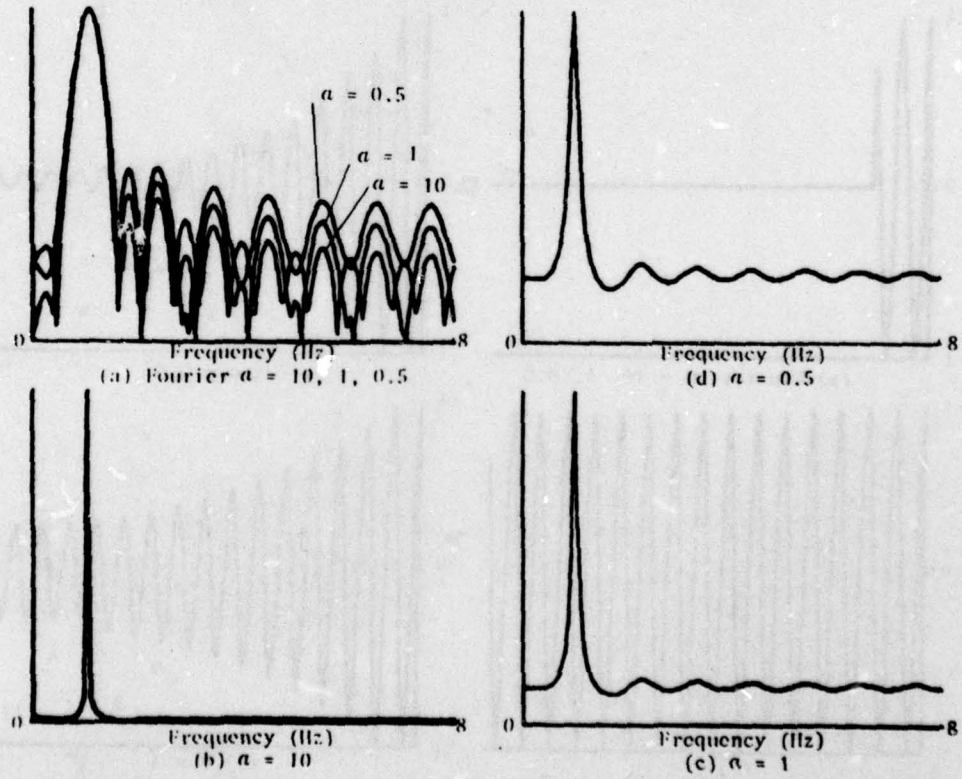


Figure 2. Fourier and Maximum Entropy spectra as a function of S.N.R.
(A: Fourier, B,C,D: Maximum Entropy)

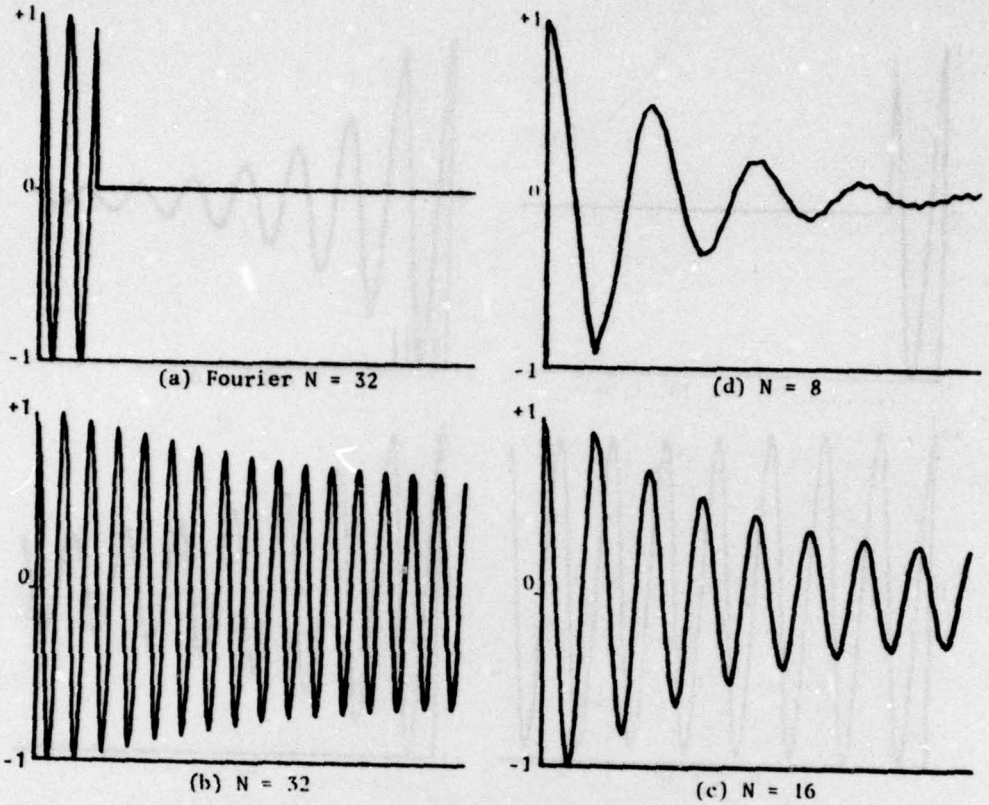


Figure 3. Extrapolation of sine wave as a function of number of lags
(A: Fourier, B,C,D: Maximum Entropy)

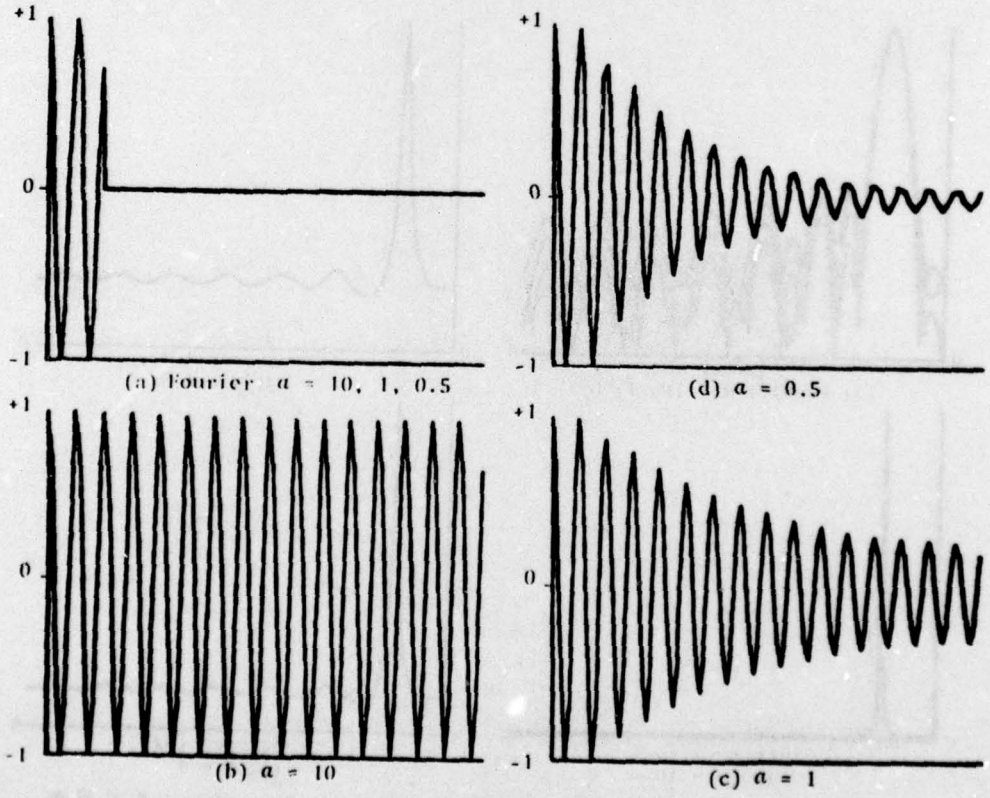


Figure 4. Extrapolation of sine wave in white noise for differing S.N.R.'s
(A: Fourier, B,C,D: Maximum Entropy)

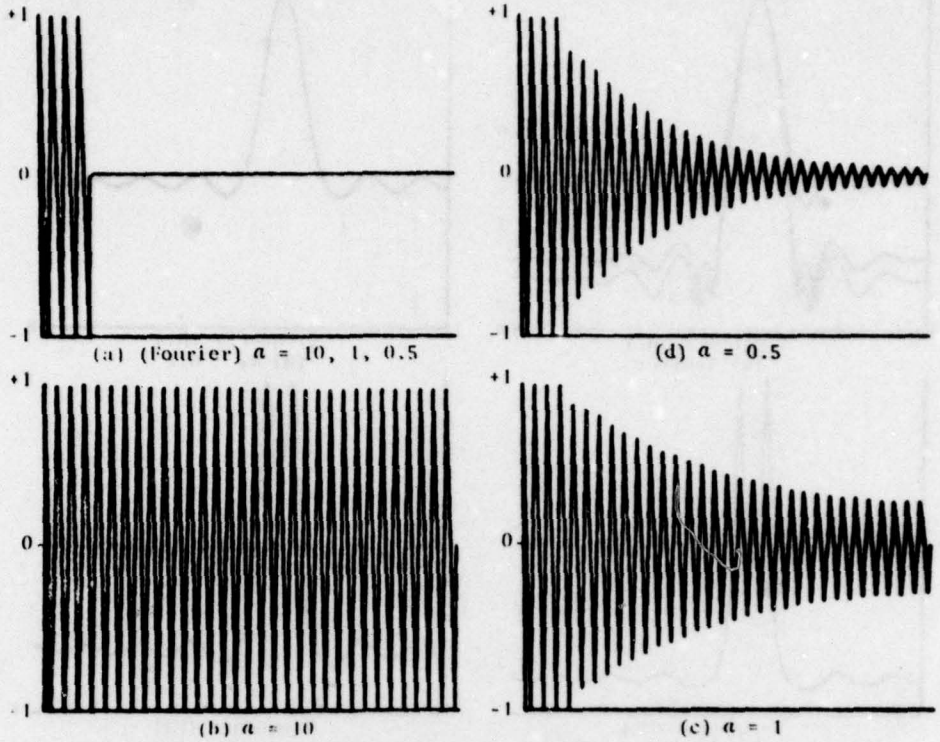


Figure 5. Extrapolation of sine wave in white noise for differing S.N.R.'s (A: Fourier, B,C,D: Maximum Entropy)

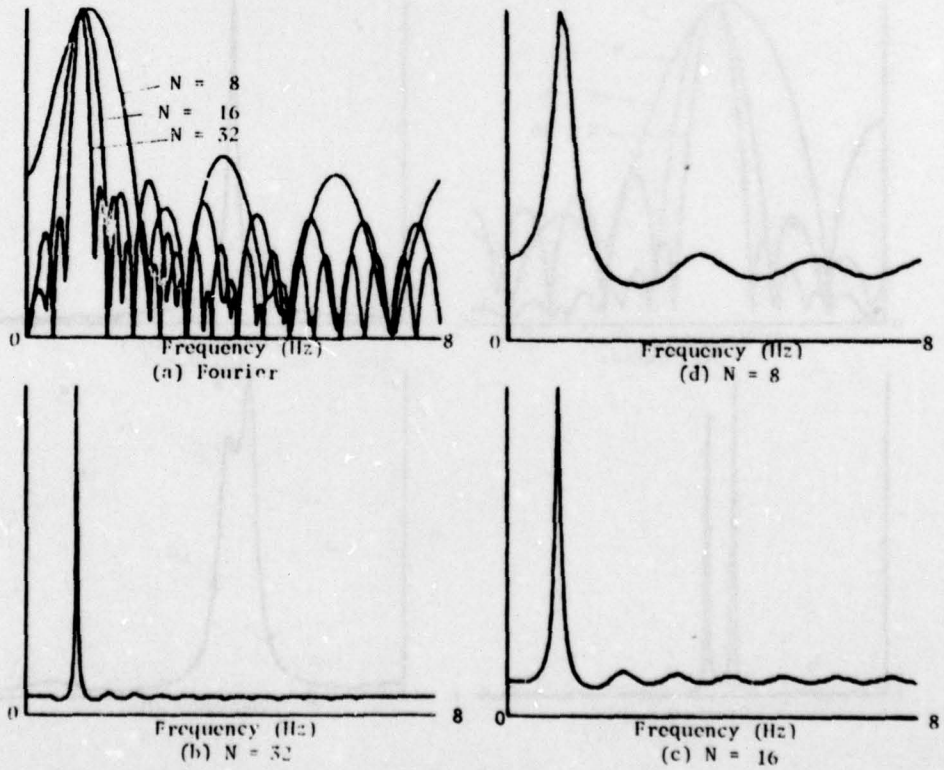


Figure 6. Fourier and Maximum Entropy spectra as a function of number of lags (A: Fourier, B,C,D: Maximum Entropy)

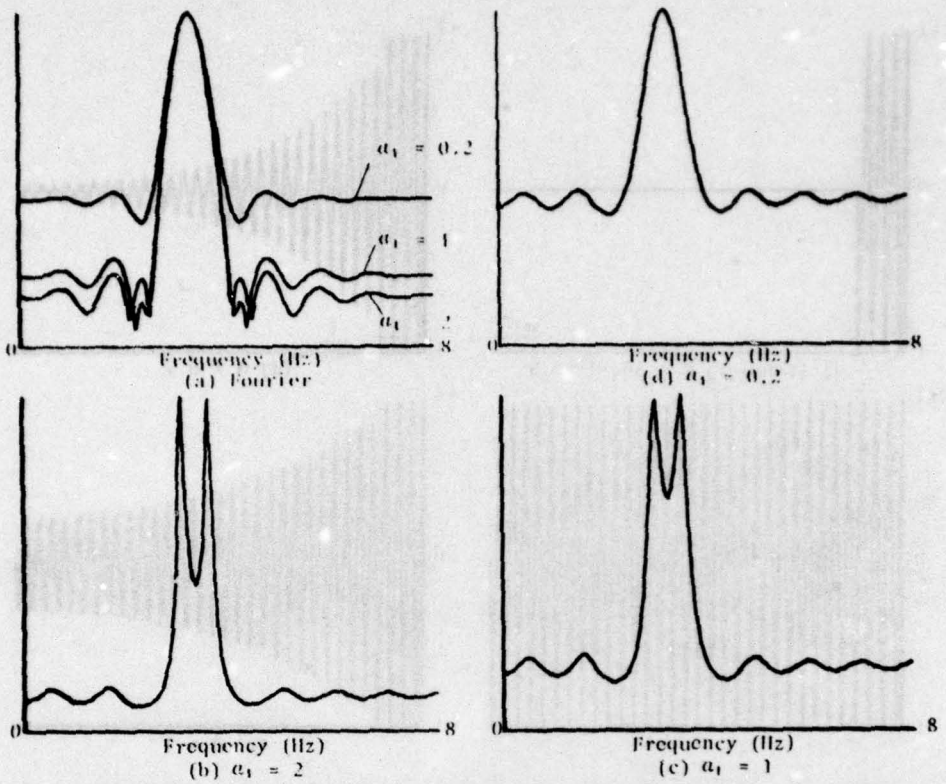


Figure 7. Resolution of two equal amplitude ($a_1 = a_2$) sine waves as a function of S.N.R. (A: Fourier, B,C,D: Maximum Entropy)

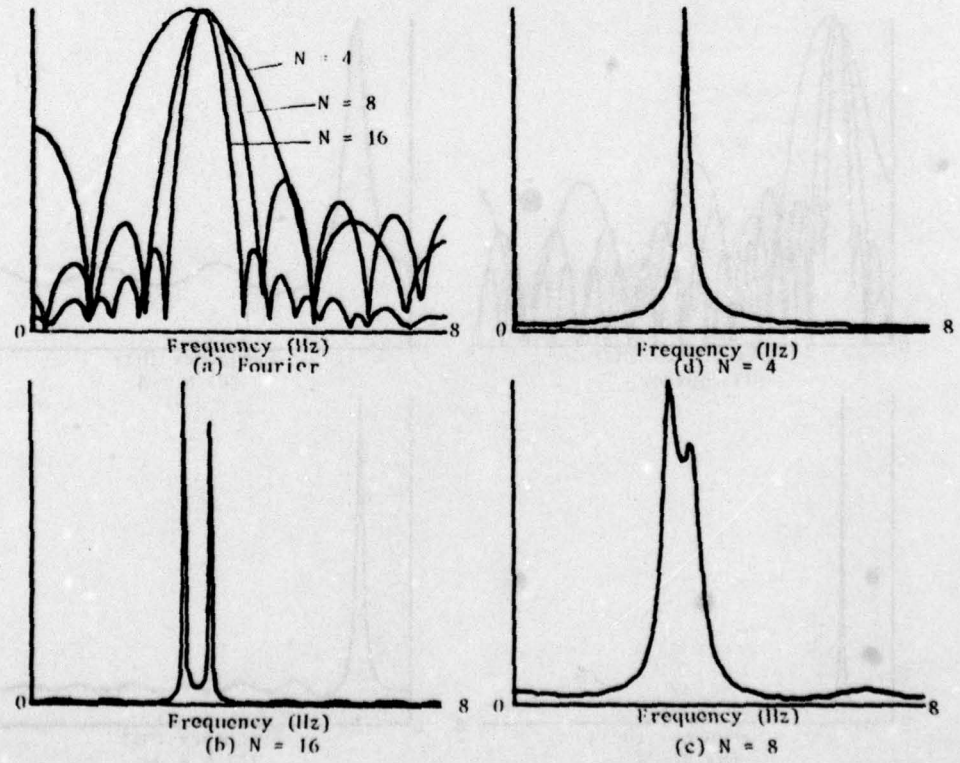


Figure 8. Resolution of two equal amplitude sine waves as a function of number of correlation lags (A: Fourier, B,C,D: Maximum Entropy)

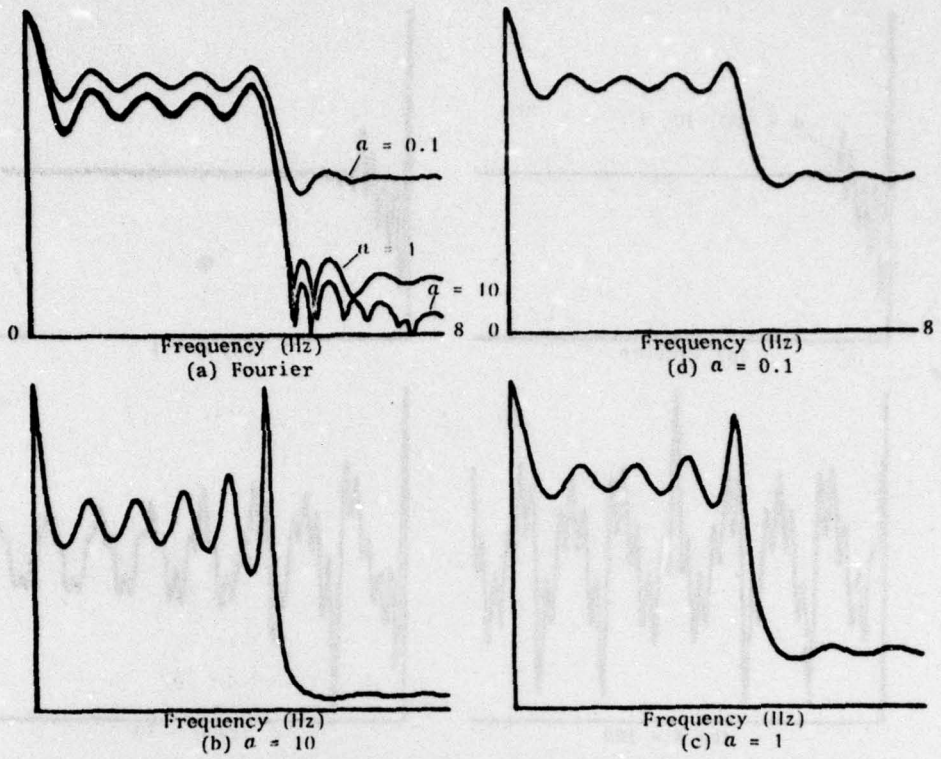


Figure 9. Fourier and Maximum Entropy spectra of 10 equi-amplitude sine waves in white noise for differing S.N.R.'s (A: Fourier, B,C,D: Maximum Entropy)

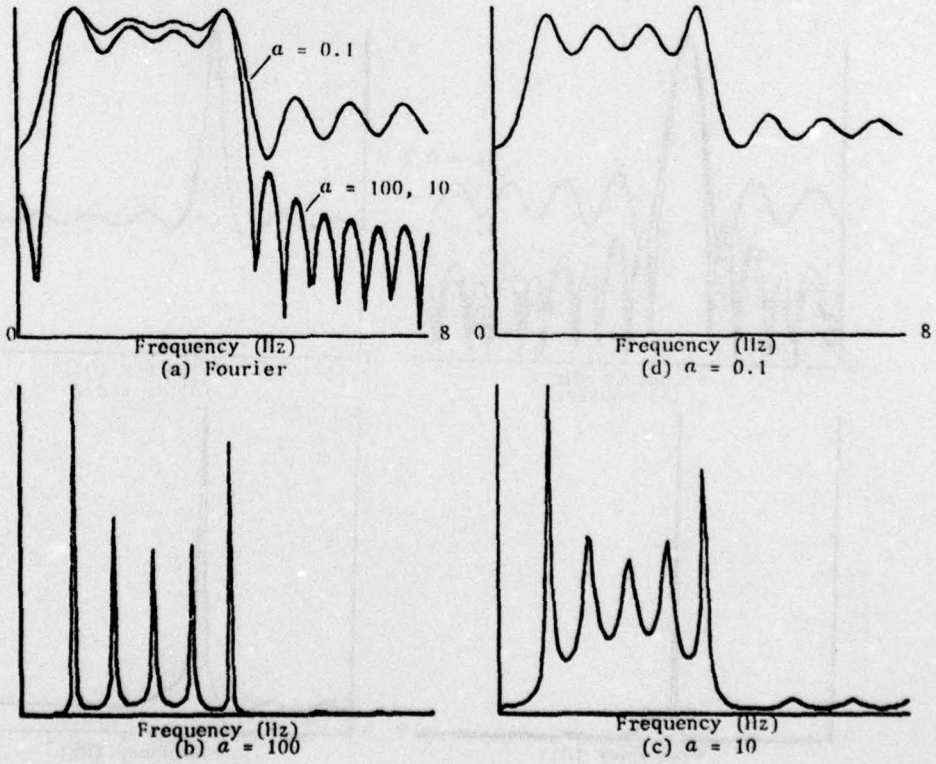


Figure 10. Fourier and Maximum Entropy spectra of 5 sine waves in white noise for differing S.N.R.'s. (A: Fourier, B,C,D: Maximum Entropy)

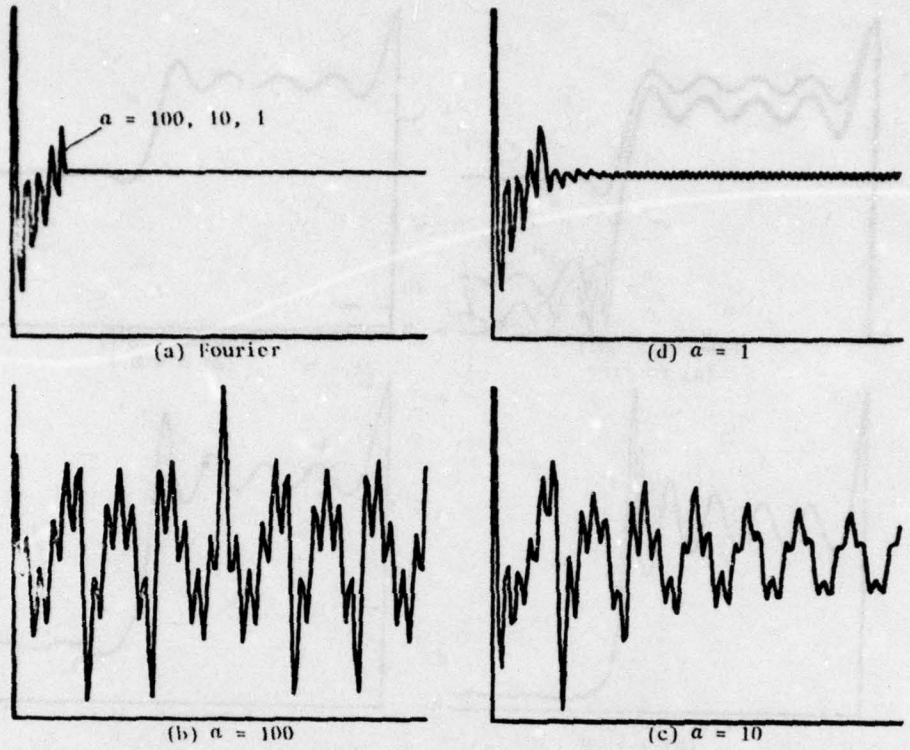


Figure 11. The autocorrelation functions corresponding to figure 12

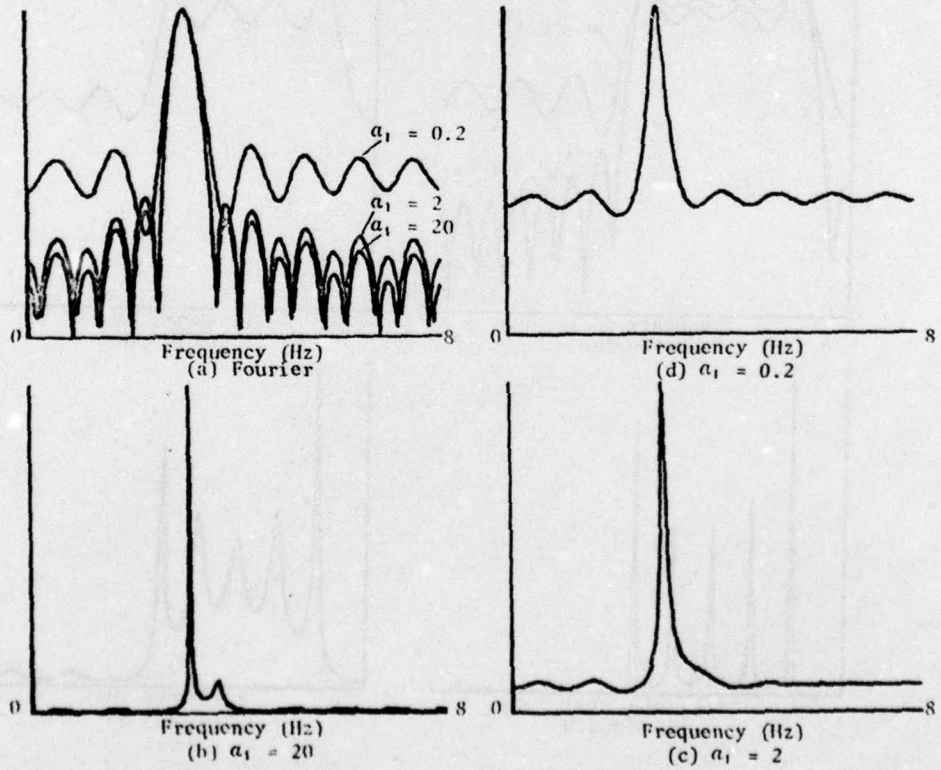


Figure 12. Resolution of two sine waves (10 dB difference in power) as a function of S.N.R. (A: Fourier, B,C,D: Maximum Entropy)

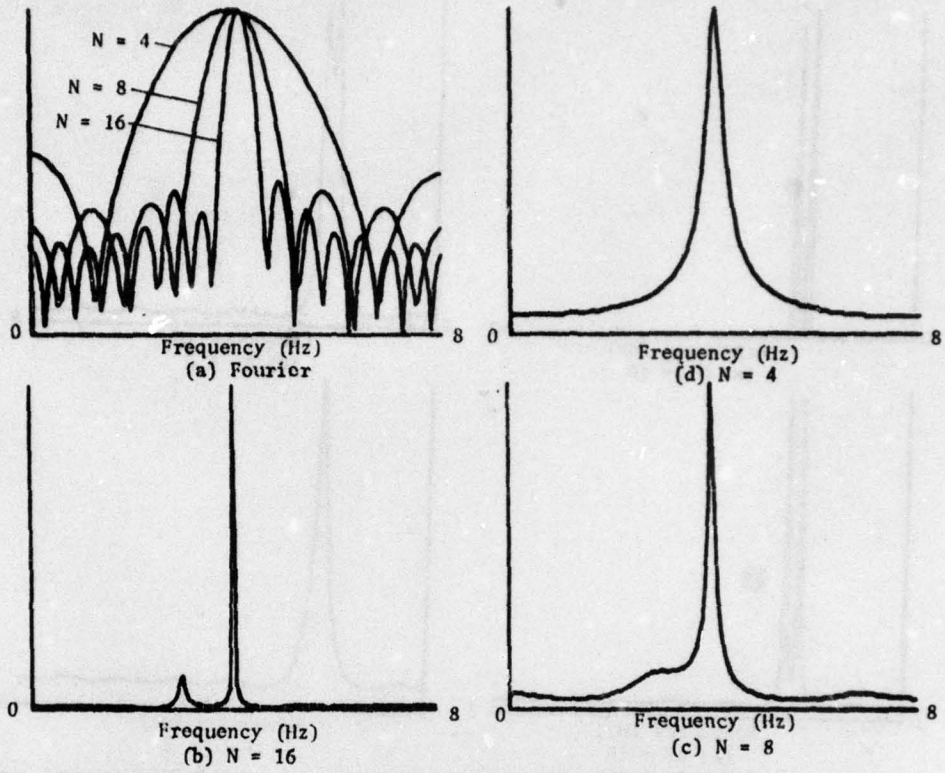


Figure 13. Resolution of two sine waves (10 dB difference in power) as a function of number of lags (A: Fourier, B,D,C: Maximum Entropy)

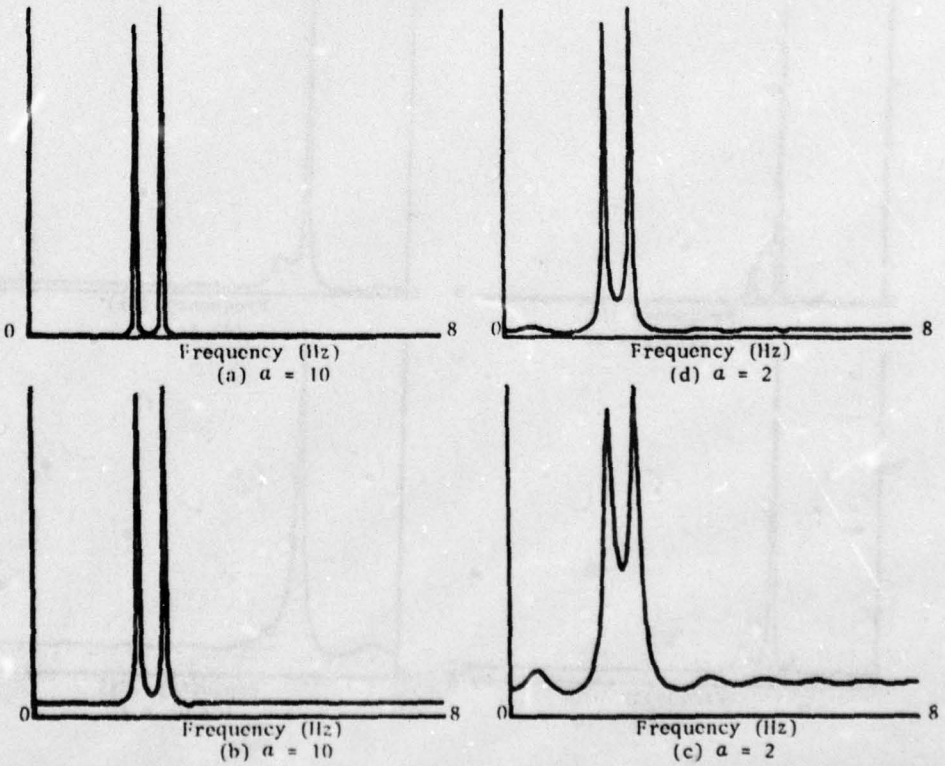


Figure 14. Linearity of Maximum Entropy Spectrum for two equal amplitude sine waves as a function of S.N.R.

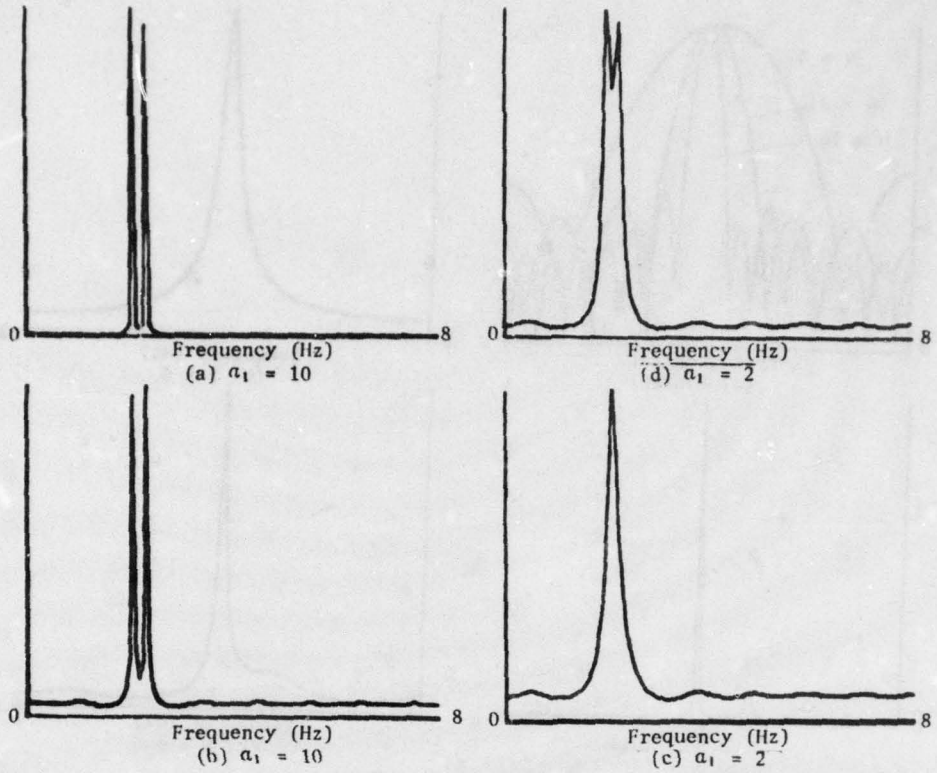


Figure 15. Linearity of Maximum Entropy Spectrum for two equal amplitude sine waves as a function of S.N.R.

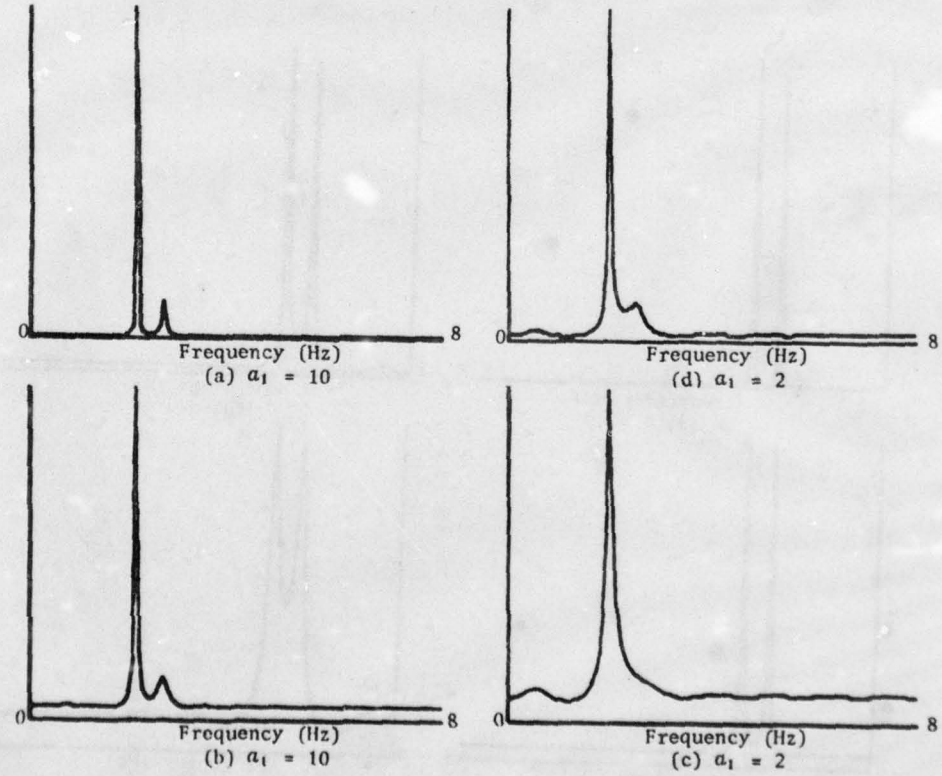


Figure 16. Linearity of Maximum Entropy Spectrum of two sine waves (10 dB difference in power) as a function of S.N.R.

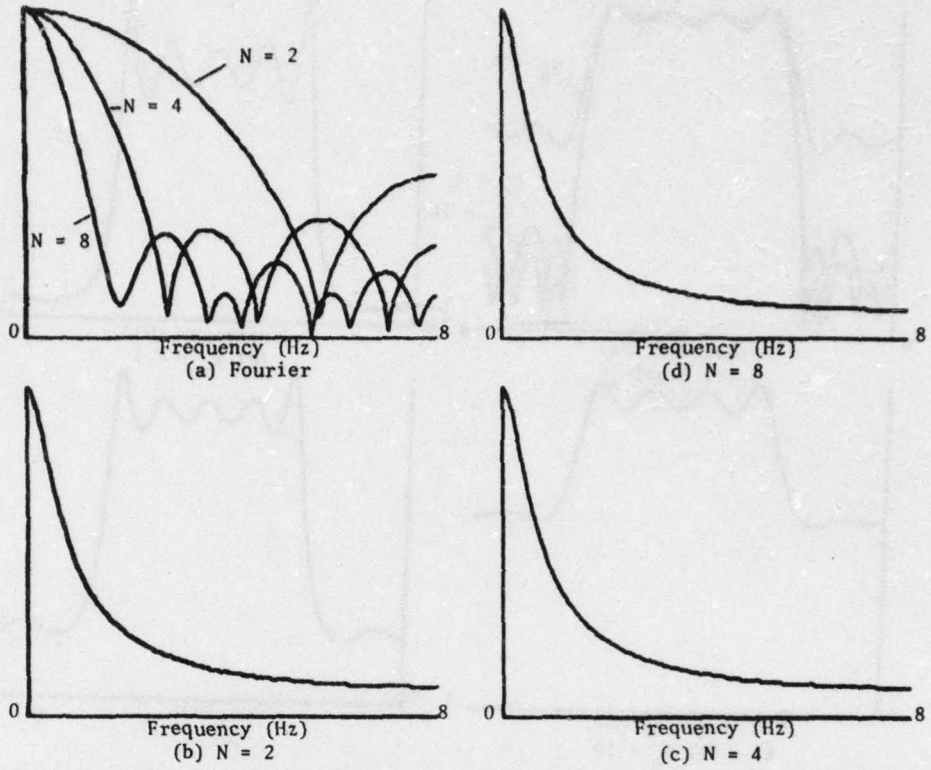


Figure 17. Fourier and Maximum Entropy estimates of Butterworth spectra as a function of N; the number of lags (A: Fourier, B,C,D: Maximum Entropy)

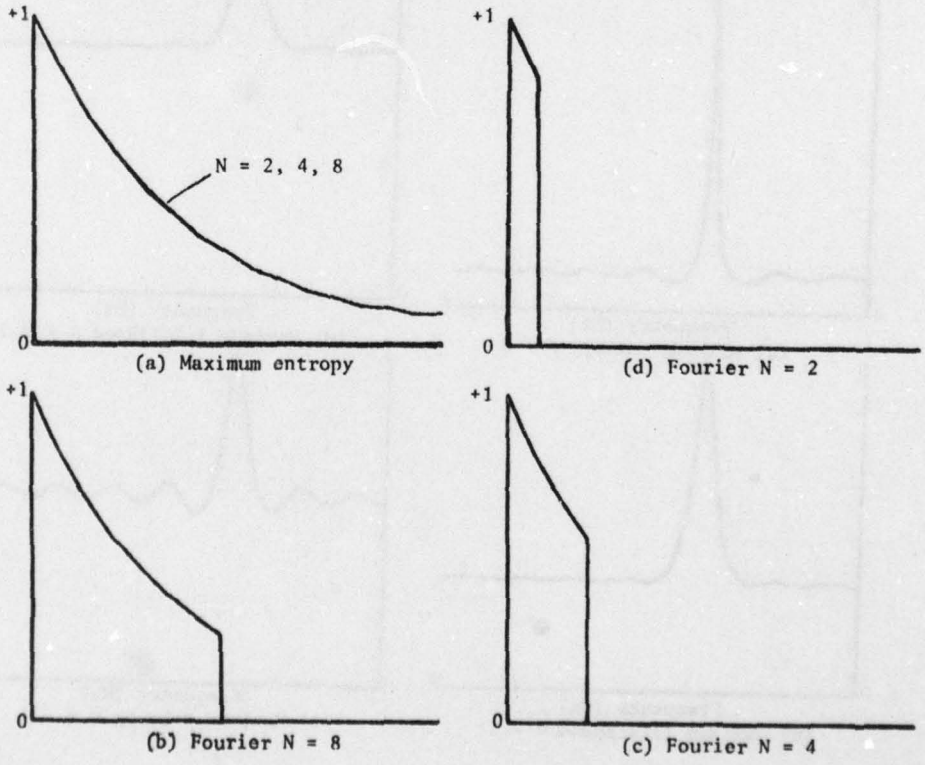


Figure 18. Autocorrelation function corresponding to figure 17 (A: Maximum Entropy, B,C,D: Fourier)

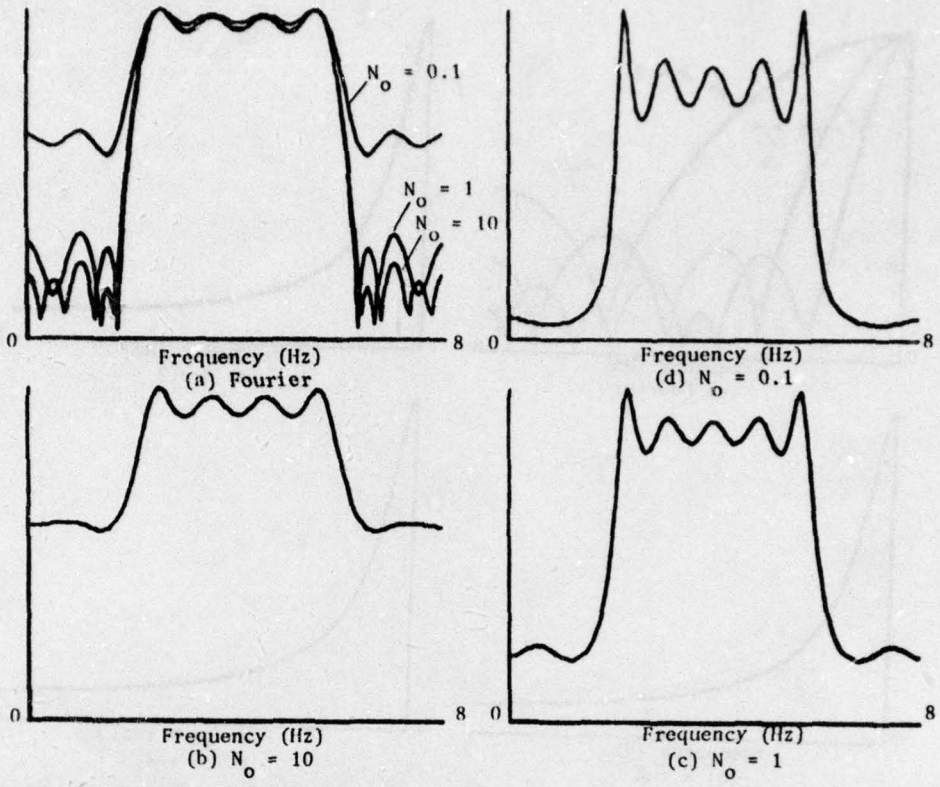


Figure 19. Fourier and Maximum Entropy spectra for band-limited white noise

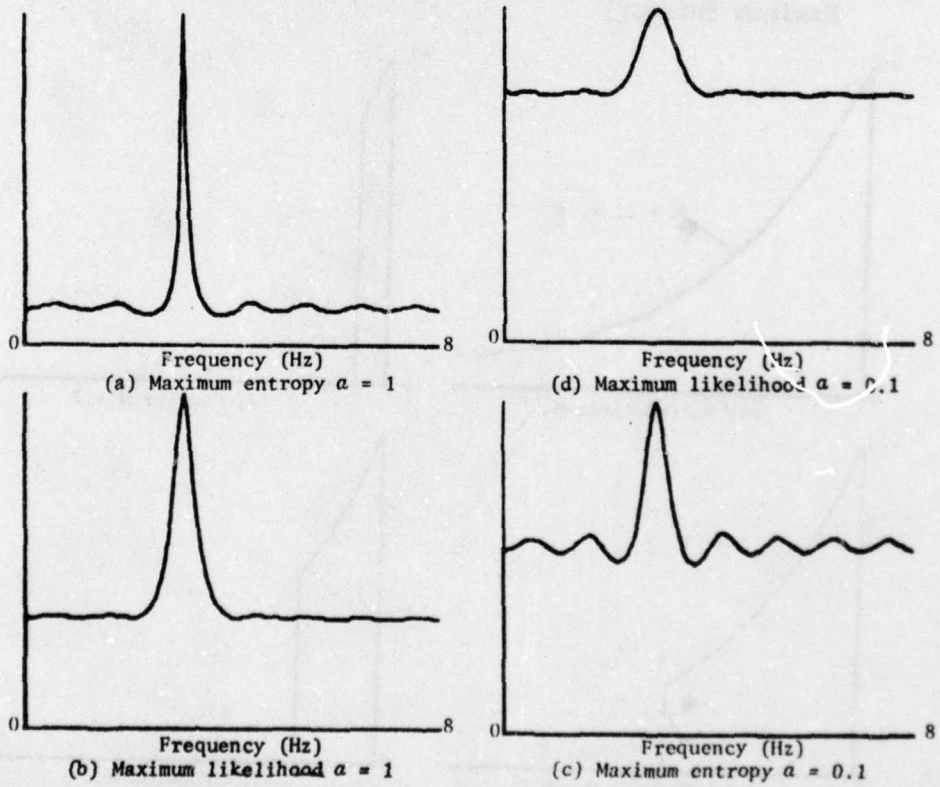


Figure 20. Maximum Entropy and Maximum Likelihood spectra for sine wave in white noise as a function of S.N.R.

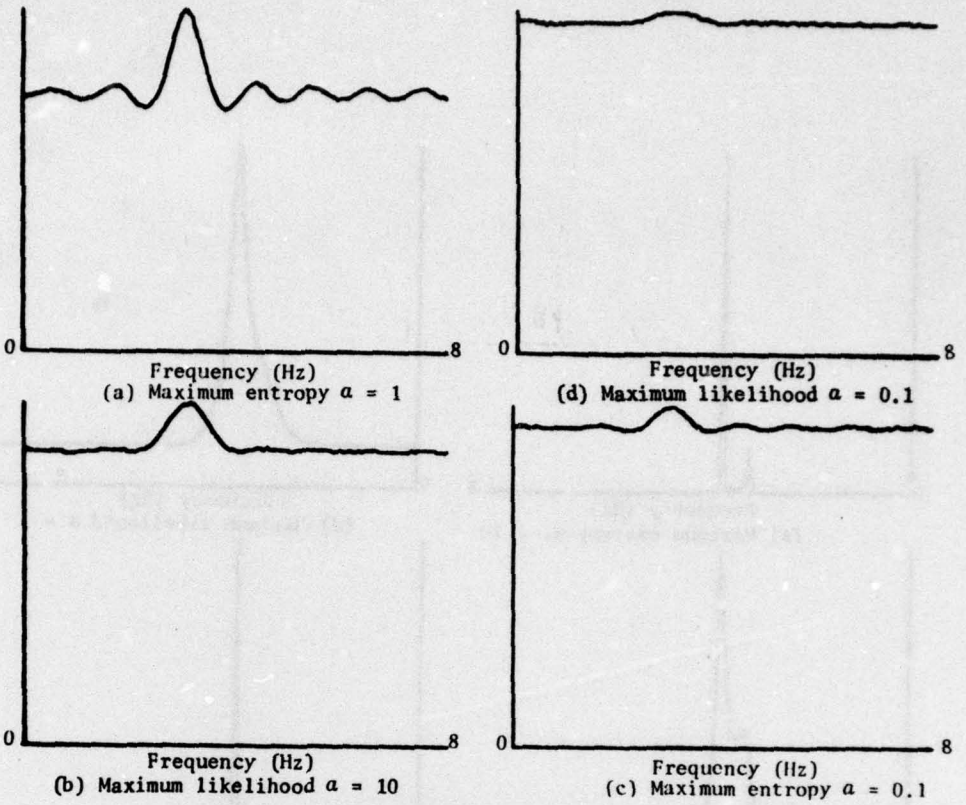


Figure 21. Maximum Entropy and Maximum Likelihood spectra for sine wave in white noise as a function of S.N.R.

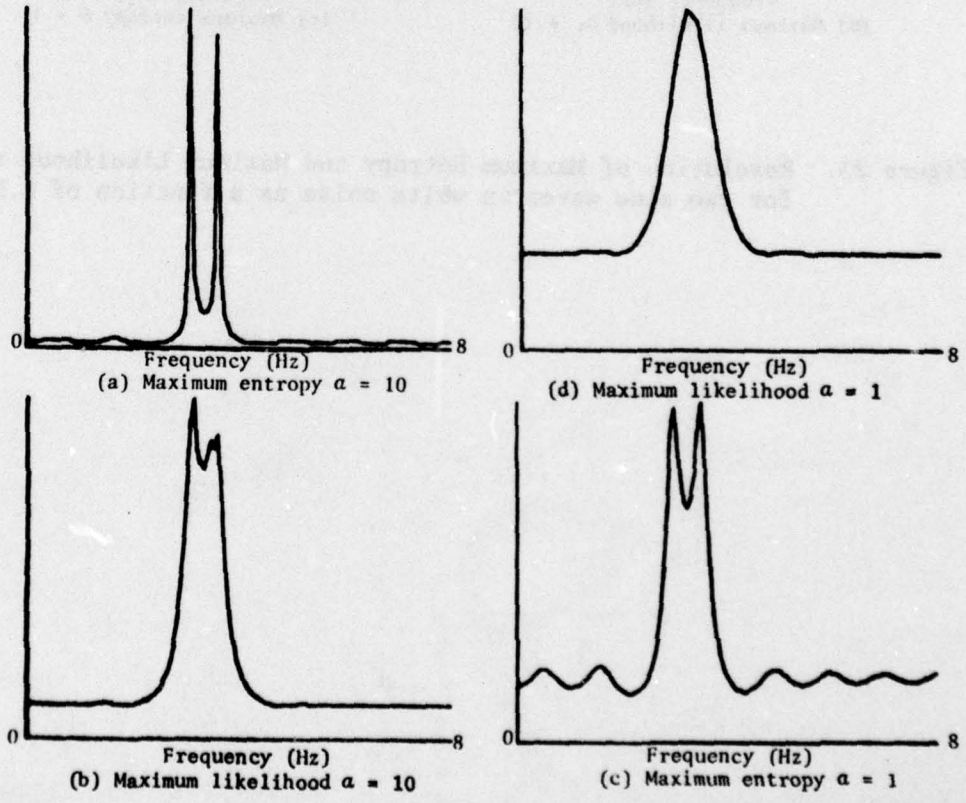


Figure 22. Resolution of Maximum Entropy and Maximum Likelihood spectra for two sine waves in white noise as a function of S.N.R.

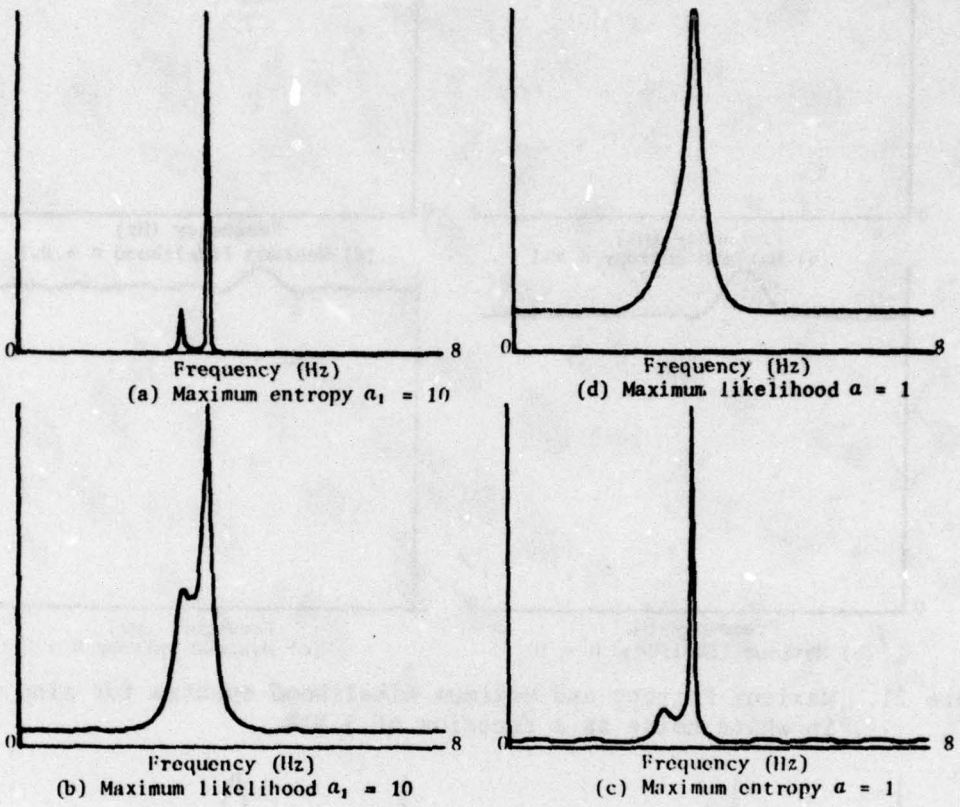


Figure 23. Resolution of Maximum Entropy and Maximum Likelihood spectra for two sine waves in white noise as a function of S.N.R.

DISTRIBUTION

EXTERNAL	Copy No.
In United Kingdom	
Defence Scientific and Technical Representative, London	1
Professor O.L. Stocklin, Loughborough Unibersity of Technology U.K.	2
Dr A. Fenwick, MASD, Camberley, Surrey, U.K.	3
In United States	
Counsellor, Defence Science, Washington	4
Professor T.G. Birdsall, University of Michigan	5
In Australia	
Chief Defence Scientist	6
Executive Controller, Australian Defence Scientific Service	7
Superintendent, Defence Science Administration Division	8
Defence Information Services Branch (for microfilming)	9
Defence Library, Campbell Park	10
Library, Aeronautical Research Laboratories	11
Library, Materials Research Laboratories	12
Defence Information Services Branch for:	
United Kingdom, Ministry of Defence, Defence Research Information Centre (DRIC)	13
United States, Department of Defense, Defense Documentation Center	14 - 25
Canada, Department of National Defence, Defence Science Information Service	26
New Zealand, Department of Defence	27
Australian National Library	28
Director, Joint Intelligence Organisation (DDSTI)	29
Superintendent, RAN Research Laboratory, Sydney	30
Professor R. Keats, Mathematics Department, University of Newcastle	31
Professor R.E. Bogner, Electrical Engineering Department University of Adelaide	32
Professor G. Newstead, Australian National University, Canberra	33
Mr J. Murphy, Australian Telecommunication Commission Research Department	34
Dr J.P. Wild, CSIRO Division of Radio Physics, Epping, N.S.W.	35

	Copy No.
INTERNAL	
Director	36
Chief Superintendent, Weapons Research and Development Wing	37
Chief Superintendent, Applied Physisc Wing	38
Superintendent, Propulsion and Marine Physics Division	39
Superintendent, Electronics Division	40
Superintendent, Systems Analysis Division	41
Senior Principal Research Scientist, Marine Studies	42
Senior Principal Resarch Scientist, Radar	43
Senior Principal Research Scientist, Electronic Warfare	44
Principal Officer, Marine Physics Group	45
Principal Officer, Underwater Detection Group	46
Principal Officer, Jindalee Project Group	47
Principal Officer, Ionospheric Studies Group	48
Mr G.C. Mountford, Underwater Detection Group	49
Dr D.H. Nichol, Underwater Detection Group	50
Dr A.K. Steele, Underwater Detection Group	51
Mr D.R. Sweet, Underwater Detection Group	52
Dr D.V. Wyllie, Marine Physics Group	53
Dr R.K. Jarrott, Jindalee Project Group	54
Dr R. Clarke, Ionospheric Studies Group	55
Author	56 - 58
P.M.D. Library	59 - 60
W.R.E. Library	61 - 62
Spares	63 - 68